Ph.D. DISSERTATION

Gated-Diode Memory Cell and Array Utilizing GIDL Current

GIDL 전류를 이용한 게이티드 다이오드 메모리 셀 및 어레이

BY

JUWAN LEE

February 2014

DEPARTMENT OF ELECTRICAL ENGINEERING AND
COMPUTER SCIENCE
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

Gated-Diode Memory Cell and Array Utilizing GIDL Current

GIDL 전류를 이용한 게이티드 다이오드 메모리 셀 및 어레이

지도교수 이 종 호

이 논문을 공학박사 학위논문으로 제출함

2014 년 2 월

서울대학교 대학원

전기컴퓨터공학부

이 주 완

이주완의 공학박사 학위논문을 인준함

2014 년 2 월

위 원 장 : _____ (인)

부위원장 : _____ (인)

위 원 : _____ (인)

위 원 : _____ (인)

위 원 : _____ (인)

ABSTRACT

In this dissertation, the gated-diode memory cell and array utilizing the gate-induced drain leakage (GIDL) current is proposed and investigated for ultra-high density memory device. In the gated-diode memory, there is no short-channel effect (SCE) which is the critical issue for scaling down of the cell in conventional FET type memories. In addition, the random access can be possible in array structure although the cells are connected serially like as conventional NAND flash memory array, because the n^+ diffusion region is directly connected to all cells which are connected in a bit-line (BL). In the gated-diode memory, Fowler-Nordheim (FN) tunneling is used for injection of electrons and band-to-band-tunneling (BTBT) induced hot-hole injection is used for injection of holes. To sense the cell state, the GIDL current is detected with negative gate bias condition. The GIDL current is increased and decreased by the stored electrons and holes, respectively.

Recent trend of nonvolatile memories are introduced in Chapter 1. In Chapter 2, Gated-diode memory utilizing the GIDL current is introduced. In Chapter 3, fabrication process and measured data of gated-diode memory cell and array utilizing GIDL current is presented. Three methods of fabrication process is represented and key process technologies are shown with SEM images. In Chapter 4, the comparison between measurement and simulation is done and improved device structure with SiGe material is

investigated to increase the sensing current (GIDL current). In Appendix, the low-frequency noise (LFN) characteristics of GIDL current in MOSFETs are investigated.

Keywords: nonvolatile memory, gated-diode, gate-induced drain leakage (GIDL), band-to-band-tunneling (BTBT), Fin, SiGe, low-frequency noise (LFN)

Student number: 2009-30927

CONTENTS

Chapter 1

Introduction

1.1 Flash Memory Technology

Flash memory is the representative MOS non-volatile memory device. The semiconductor memory devices are classified into two categories with volatile and non-volatile as shown in Fig. 1.1. The volatile memory is the random access memory (RAM) which requires a constant voltage supply to maintain the stored data. Otherwise the stored data is lost. There are two main categories in volatile memory: static RAM (SRAM) and dynamic RAM (DRAM). SRAM is often used as cache memory for the CPU thanks to their high operation speed, but is generally most expensive to produce due to its large size with six-transistors. DRAM stores a bit of data using a transistor and a capacitor; the capacitor holds a low or high voltage state (0 or 1) and the transistor acts as a switch that lets the state of capacitor to be changed. Since capacitors do not hold a charge indefinitely, DRAM cells must be frequently recharged (refreshed) to

avoid losing the data.

The non-volatile memory is read only memory (ROM) which the memory data is defined during manufacturing and not modifiable. The non-volatile memory includes programmable read only memory (PROM) and mask read only memory (Mask ROM). And PROM includes electrically programmable read-only memory (EPROM) and electrically erasable programmable read-only memory (EEPROM). EPRM is erased by exposure to UV light and programmed electrically. And EEPRM can be both erased and programmed electrically. Flash memory is a kind of EEPROM where the erase procedure of entire chip or sub-array within the chip is performed at one time [1], [2]. Due to the erase operation being much faster than the prior EPROM or EEPROM devices, these devices came to be called flash memory. Moreover flash memory has high density and high scalability. So, the demand for flash memory device increased rapidly in the electronic devices such as cellular phone, PDA, mp3 player, digital still camera, USB drive, digital camcorder, solid state drives (SSDs) and others as shown in Fig. 1.2.

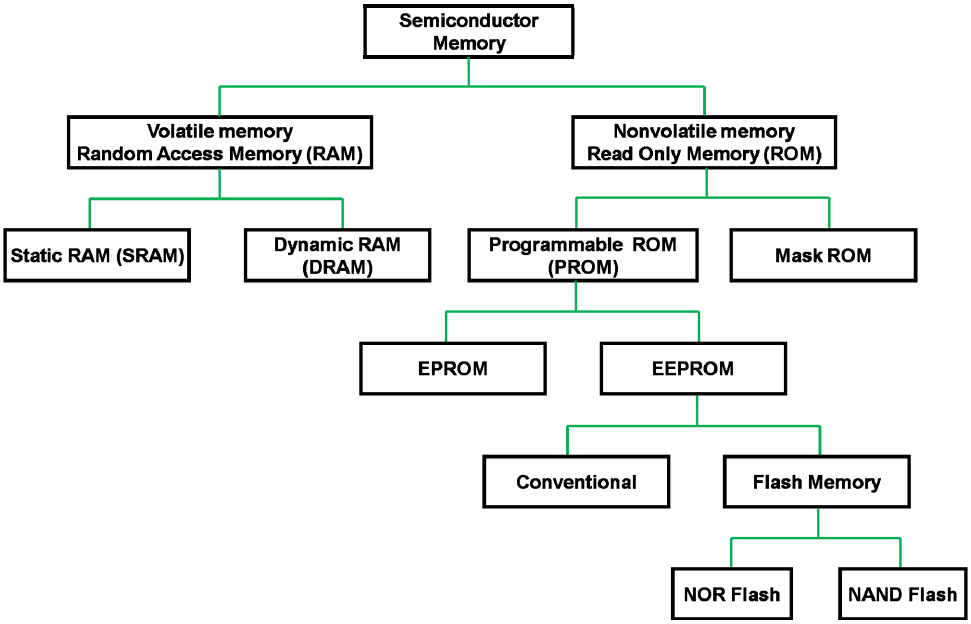


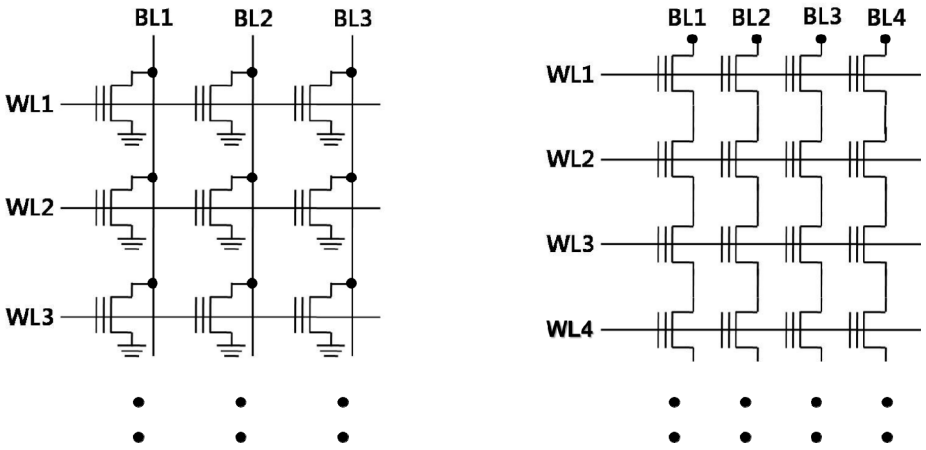
Fig. 1.1. Classification of semiconductor MOS memories.

Fig. 1.2. Applications of NAND flash memory.

1.2 Basic Operation of Flash Memory

Flash memory is divided into two types according to their array organization. One is the NOR-type flash memory which enables the fast random access operation during the read operation. So, it is mainly used for code storage, where the operating system is stored and is executed by the microprocessor. But bit-line contact is needed for every cell to access to every cells randomly. This causes relatively large cell area about $10F^2$ where F is minimum feature size. Moreover, it is impossible to program a large amount of cells at the same time because of the large power consumption from the low injection efficiency ($\sim 1\times10^{-6}$) of channel hot electron injection (CHEI) mechanism [3] which is used for programming in NOR flash memory [4], [5].

The other type is NAND flash memory which is suitable for high-density mass storage applications. The memory cells in NAND flash memory are connected in series and only one bit-line contact is needed in a string array architecture as shown in Fig. 1.3. Therfore the effective cell size is much smaller than NOR flash memory, and then, smaller chip size and lower cost per bit can be possible. Moreover, NAND flash memory exhibits fast page writes due to the ability to write 4-8kB simultaneously resulting in very high sequential write throughput [3].



(a) (b)

Fig. 1.3. Array architecture of (a) NOR and (b) NAND flash memory.

The memory operation in flash memory is performed by the charge transport through the gate dielectric and chare storage at the floating gate or charge trap layer. Fowler-Nordheim (FN) tunneling is usually used for charge injection to the charge storage layer in NAND flash memory. The FN tunneling occurs at sufficiently high gate voltage. Electrons in the conduction band of the channel can tunnel through the gate dielectric and stored at the charge storage layer. These stored electrons increases the threshold voltage of flash memory device. And the threshold voltage can be decreased by introducing holes into or removing electron from the storage layer. The threshold voltages are switched between two distinct values, logical “1” (erased state) and the logical “0” (written or programmed state).

1.3 3D stacked NAND Flash Memory

The scaling down of the device in NAND flash memory is the most important issue for low cost and high-density properties. However, there are several problems to decrease the cell size [6]-[10]. The main problem is the limitation of downscaling of line width with photo lithography process. In order to make up for the limitations, various technologies such as optical proximity correction (OPC), phase shift mask (PSM), and double patterning technology (DPT) are used. However the gap will be widened beyond 20 nm technology node until we find new solutions. Until that time, it will be difficult to find out the appropriate solutions for compensating the gap in further shrinking.

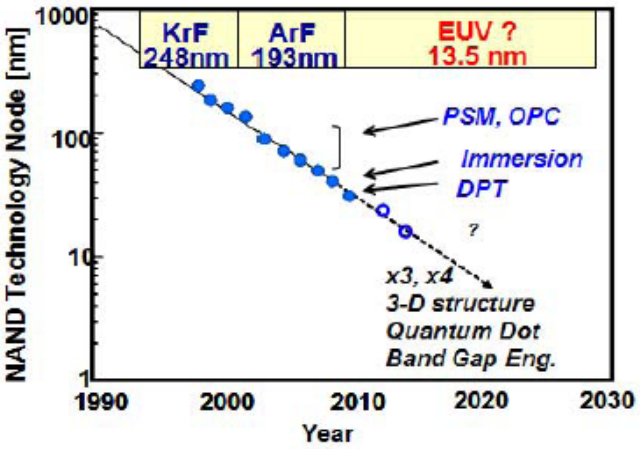
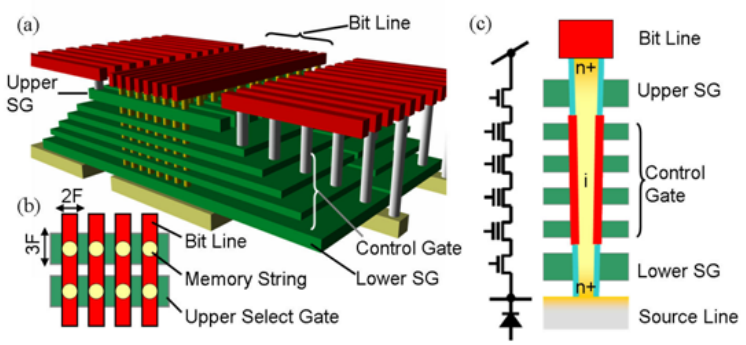


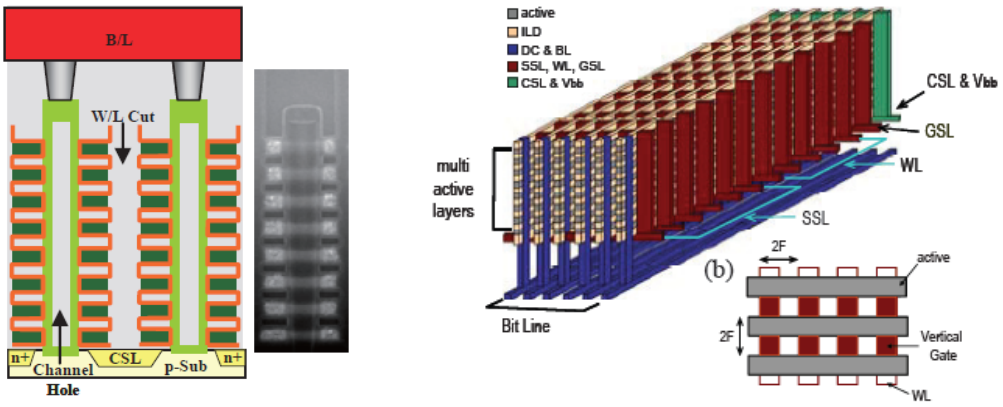
Fig. 1.4. The limitations of photo lithography in NAND flash memory technology nodes.

As the gate length of transistor shrinks, it is necessary to increase the doping concentration of substrate so that we can keep the effective channel length preventing S/D punch through. In NAND Flash, however, high doping concentration can deteriorate boosted potential at transistor channel during program-inhibit condition at unselected BL, which narrows V_{pass} bias operating window [11]. To overcome these problems, various 3-dimensional (3D) stacked NAND flash memory architentures have been introduced as shown in Fig. 1.5. Bit Cost Scalable (BiCS) NAND flash memory is introduced to increase the layers of 3D stacked cells without increase of cost by Toshiba Corp of Japan in 2007 [12]. Cell size of BiCS memory is $6F^2/n$, where n is the number of control gate plates. A single cell is accessed by the control gate on the string which is selected by a bit line and a row select line. The bottom of memory plug is connected to source diffusion formed on the silicon substrate. For the erase operation of BiCS flash memory, hole current which is generated by gate induced drain leakage (GIDL) near the lower select gate is used. In 2009, Samsung Electronics introduces Tera-bit cell array transistor (TCAT) [13] with gate replacement process. Thanks to the gate replacement process, metal-oxide-nitride-oxide-silicon (MONOS) structure can be obtained. And Vertical Gate (VG) NAND flash memory which is channel stack type 3D NAND flash memory is introduced by Samsung Electronics [14]. VG-NAND flash includes WL, BL, CSL, horizontal active string with pattern, VG (for SSL, WL, GSL), charge trap layers between active and vertical gate, vertical plugs of DC, source and active body. Cell size of VG-NAND is $4F^2$ per layer and the active dimension can keep getting down without scaling

issue, making the cell size below $4F^2$ per layer.



(a)



(b) (c)

Fig. 1. 5. 3-D stacked NAND flash memory (a) Bit Cost Scalable (BiCS) NAND flash memory by Toshiba. (b) Terabit Cell Array Transistor (TCAT) by Samsung Electronics. (c) Vertical Gate NAND flash memory (VG-NAND) by Samsung Electronics.

1.4 Emerging New Memory (MRAM, PRAM, RRAM)

In order to overcome the limitations of conventional memory, various alternatives as emerging new memories such as Magnetoresistive RAM (MRAM), Phase change RAM (PRAM) and Resistive RAM (RRAM) have been studied. These new memories have some advantages compared with conventional memory. First, the limitation in the scaling of the cell size is relieved because they are non-charge based memory. Also, there is no short channel effect problems according to the decreased cell size. Second, both random access and high density characteristics can be achieved at the same time thanks to the simple structure (2-terminal) and scalability of new memories. Despite such advantages, it has several issues such as unknown switching mechanism and high switching current [15], [16]. And it is difficult to accomplish high density memory with conventional structure due to limitation of photolithography [17], [18]. Moreover, the reliability properties are not well characterized yet. So, it needs more study and characterization.

Chapter 2

Gated-diode Memory utilizing GIDL Current

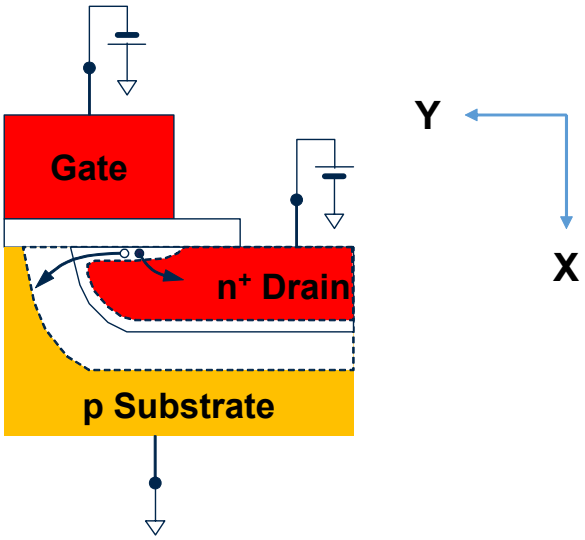
In this chapter, the new memory device with gated-diode structure is introduced for high density nonvolatile flash memory. This memory device utilizes the gate-induced drain leakage (GIDL) current not a channel current used in conventional FET type flash memories.

2.1 Introduction of GIDL Current

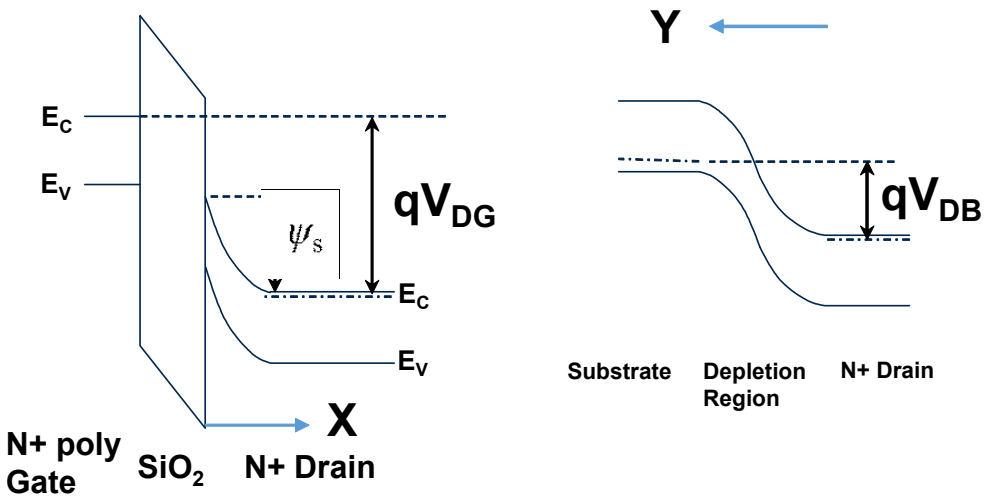
Fig. 2. 1 shows the generation mechanism of GIDL current. Fig. 2. 1 (a) shows a cross-sectional view of a n-MOSFET device. When the drain bias is high and gate is negatively biased, a depletion region is formed underneath the drain overlapped by the gate region and a high field is created in the depletion region. Electron-hole pairs are

generated by the band-to-band tunneling (BTBT) process and collected by the drain and substrate separately. In Fig. 2. 1 (b) and (c), a vertical and lateral energy band diagrams are presented. For the BTBT process, the vertical field in the drain is the dominant field. In addition, the vertical field depends on the band bending ψ_s, as shown in Fig. 2. 1 (a), and the ψ_s is strongly related to the drain doping concentration. Therefore, the drain doping concentration is also an important parameter in the GIDL current. In Fig. 2. 1 (c), the hole flows into the substrate due to the lateral field. The larger lateral field would increase the lateral momentum of the electrons, which enhances the total momentum of the electrons. In addition, a larger lateral field would reduce the energy barrier that the electrons tunnel through. This phenomenon is called the drain-induced energy-barrier reduction effect [19]. So, the GIDL current is dependent on both the drain-to-gate voltage and drain-to-substrate voltage.

The GIDL current is known as the unwanted leakage component in deep-submicrometer CMOS technology [20]. In addition, GIDL is the major leakage compoent which worsen the retention characteristics in the DRAM cell transosstors. Espacially, the DRAM cell becomes more susceptible to GIDL when the storage node stores a "1" and a negative bias is applied to the gate. The voltage drop across the gate oxide creates a vertical electric field that leads to a higher leakage current. And GIDL is also one of the major mechanism which degrades program disturbances in NAND flash operation [21].



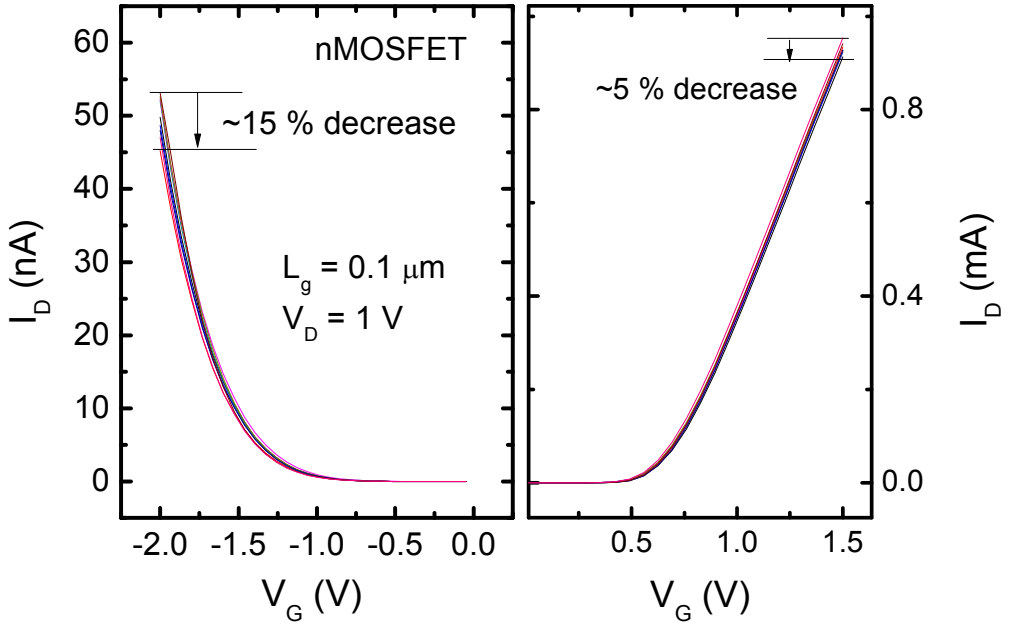
(a)



(b) (c)

Fig. 2. 1. Generation mechanism of GIDL current (a) cross-sectional view of nMOSFET (b) energy band diagram with X-direction (c) energy band diagram with Y-direction

In contrast, a capacitorless 1T-DRAM cell using GIDL current for write operation was introduced and many studies have been reported [22]. Compared with the conventional write operaion with impact ionization current, write operation with GIDL current provides low-power and high-speed operation. Furthermore, the GIDL current is used as a read method for dual-bit SONOS nonvolatile memory cells [23]. Here, the GIDL current is demonstrated to be more sensitive to charge stored locally within the gate dielectric stack, as compared with the transistor threshold voltage (V_T). Thus the sensing of GIDL rather than V_T is advantageous for dual-bit SONOS nonvolatile memory cell read operation, because it mitigates the complementary-bit disturb issue and hence facilitates gate length scaling. Ahead of utilizing the GIDL current, we investigated reliability of GIDL current in MOSFETs. First of all, we investigates the distribution of GIDL current through the whole wafer and compared with that of channel current. As shown in Fig. 2. 2, the variations of channel and GIDL currents are ~5% and ~15%, respectively. Although the variation of GIDL current is larger than that of channel current, reasonable distribution characteristic for using GIDL current as a sensing current is shown.



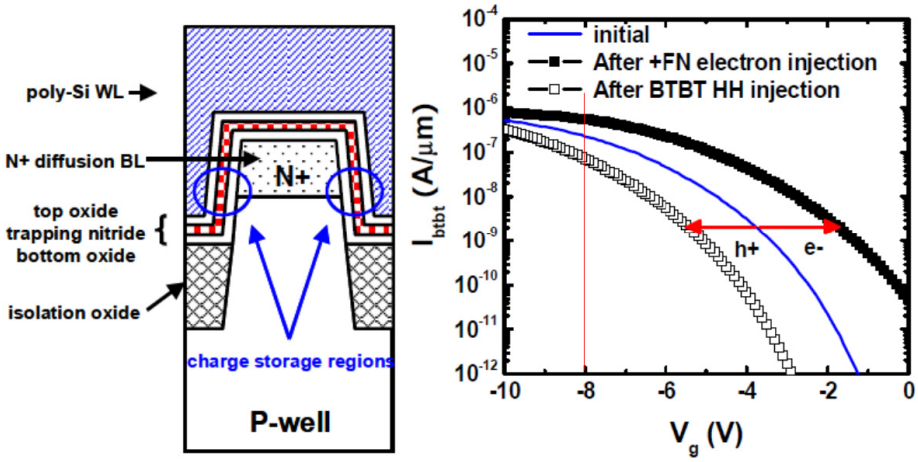
(a) (b)

Fig. 2. 2. Distribution of (a) GIDL and (b) channel currents in MOSFETs through the 12-inch whole wafer.

After that, the low-frequency noise (LFN) of GIDL current is compared with that of channel current in MOSFETs and the results are represented in appendix part. Synthetically, the LFN of GIDL currents shows much smaller than that of channel currents. So, it can obtain the advantage in LFN characteristics for using GIDL currents.

2.2 Introduction of Gated-diode Memory

The gated-diode memory device was proposed and fabricated in the previous study [24], [25]. But the sensing current window at read bias condition was not enough for memory operation and investigation of memory performance in array structure was in sufficient in the previous study. As shown in Fig. 2. 3 (b), the sensing current at high current state is only ~10 times larger than that at low current state at read bias condition of V_g = -8 V.

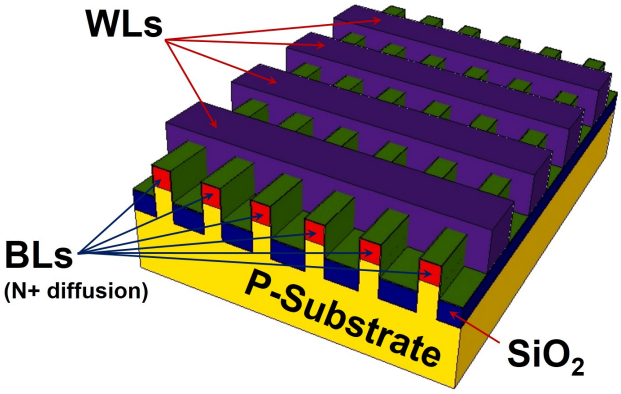


(a) (b)

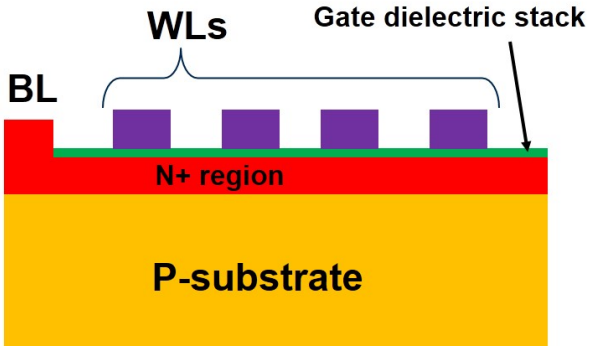
Fig. 2. 3. (a) Device structure and (b) I-V characteristics after electron or hole injection [24], [25].

Moreover, only disturbance characteristics in array structure was investigated. So, much more sensing current window at read bias condition and investigation in array structure are needed.

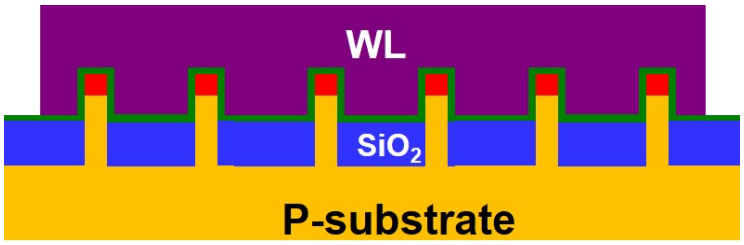
Fig. 2. 4 (a) shows schematic diagram of proposed gated-diode memory cell array, (b) shows the cut-plane with BL direction, and (c) shows the cut-plane with WL direction. Here, the gate dielectric stack is composed of tunneling oxide-charge trapping layer-blocking dielectric for memory operation. There is a critical different point compared with conventional FET type NAND flash memory cell array. It is that the n^+ diffusion region of BL is perfectly connected through the WLs even at the channel region of conventional NAND flash memory as shown in Fig. 2. 4 (b). In other words, the channel does not exist in the gated-diode memory. So, the short channel effect which is ciritical problem in scaling of conventional NAND flash memory also disappears in gated-diode memory. In addition, all the other WLs except the selected WL can be floated or grounded meanwhile they should be biased as a pass voltage to turn-on the channel in conventional NAND flash memory. So, the gated-diode memory acts as two-terminal memory devices and can have the merits of those although they are three-terminal device (WL, BL, substrate). As a result of that, the source contact is not needed.



(a)



(b)



(c)

Fig. 2. 4. (a) Schematic diagram of proposed gated-diode memory cell array, (b) cut-plane with BL direction, and (c) cut-plane with WL direction.

Fig.2. 5 shows the circuit diagram of cell array based on proposed gated-diode memory. Compared with the conventional NOR and NAND type flash memories, the merits of both can be achieved. High density property of NAND type can be achieved thanks to the serial connection of cells. And fast random access property of NOR type can be also achived thanks to the direct connection of n^+ diffusion region to the all cells which is connected serially in a BL. And the p-type body is all connected to the common body line (CBL).

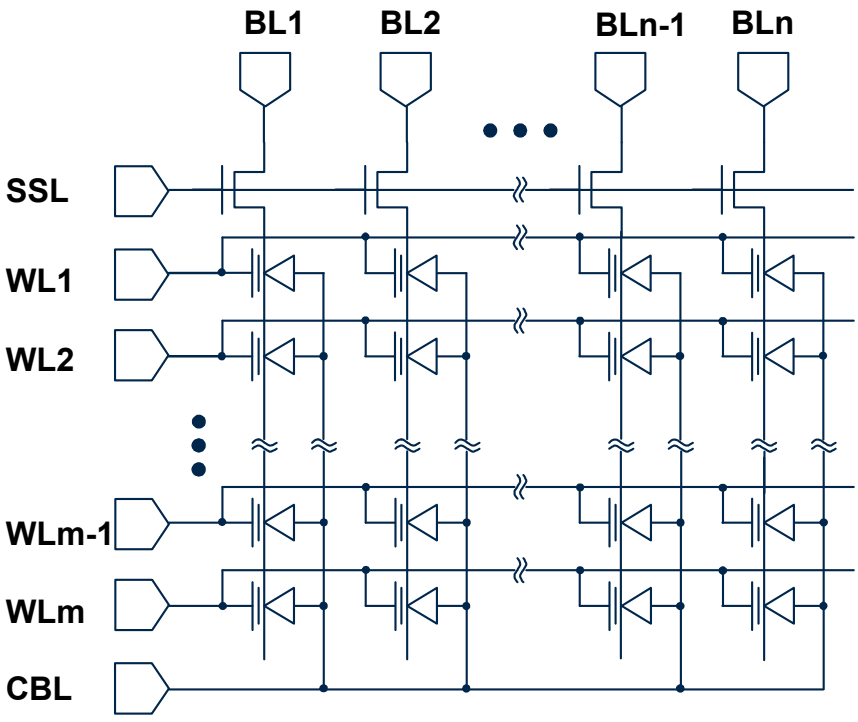


Fig. 2. 5. Circuit diagram of cell array based on proposed gated-diode memory

The fin structure is used to enable overapping of the gate with p-n^+ junction at the sides of fins as shown in Fig. 2. 4 (c).

In the proposed gated-diode memory, the programming is performed with FN tunneling and erasing is performed with hot-hole injection mechanism. As shown in Fig. 2. 6, electrons are stored in the charge trapping layer after programming performance. And these electrons enhance the vertical electric field, then, increase the GIDL current (on-state). That is, the read current is high at the programmed state meanwhile the read current is low at the programmed state in conventional NAND flash memory. And holes in the charge trapping layer after erasing performance decrease the vertical electric field, then, decrease the read current (off-state). And corresponding energy band diagrams are also represented. As shown in the band structure, band bending at the surface of n^+ diffusion region in programmed state is much larger than that in erased state at the same V_{DG}. So, the GIDL current at programmed state is much larger than that at erased state. As a result, the memory performance can be achieved by sensing the current difference from these two states.

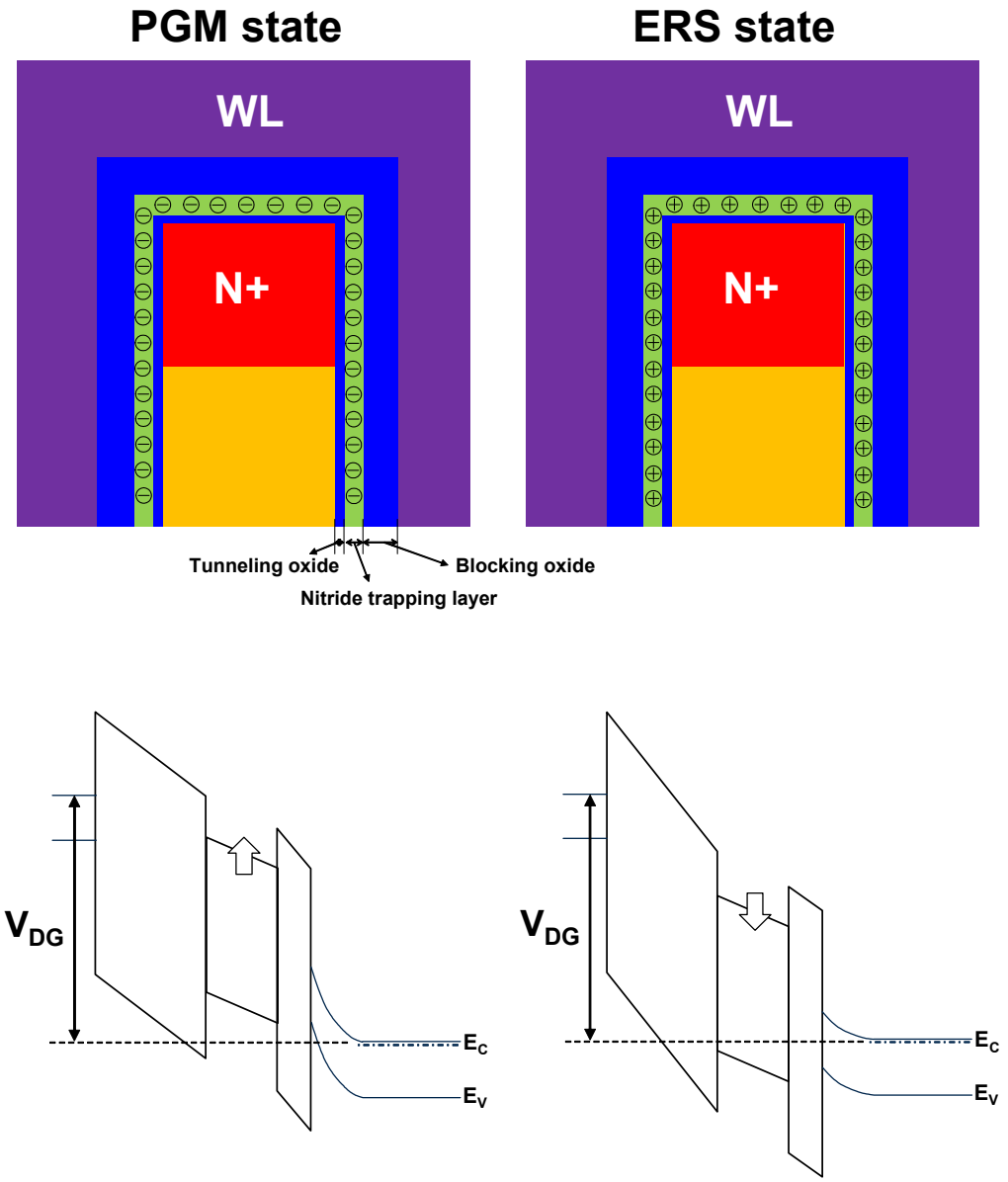


Fig. 2. 6. Schematic view which shows the programmed and erased state with the stored charge at nitride charge trapping layer.

Chapter 3

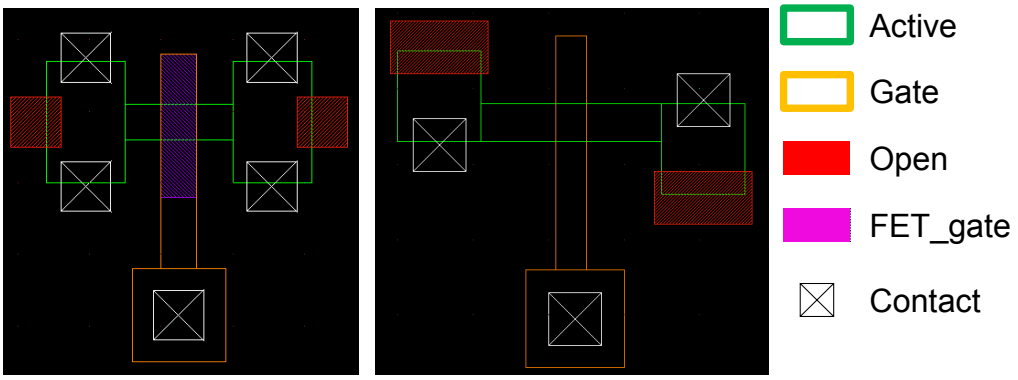
Fabrication and Electrical Characteristics of SONOS Gated-diode Memory Cell and Array utilizing GIDL Current

In this chapter, fabrication process and measured data of SONOS gated-diode memory cell and array are presented. The fin structure is adopted to implement the proposed Gated-diode memory cell & array. And the fabricated gated-diode memory cell & array show good memory performance.

3.1 Fabrication Process of Gated-diode Memory with Fin Structure

To verify the memory operation, the SONOS gated-diode memory cell and array are fabricated. Fig. 3. 1 shows the mask layout of fabricated (a) conventional FET type and

(b) gated-diode memory cell. Six mask layers (Active, Gate, Open, FET_gate, Contact, Metal pad) are used for fabrication. The nitride side-wall spacer is used to define the active fin in order to obtain narrow fin width (W_{Fin} < 100 nm). So, the active fin is formed at the border of Active mask layer. Then, two fins at both side of single Active layer are across with the gate. In order to make these two active fins operate independently, the fin is disconnected by etching the fin at the region of Open layer. Therefore, two cells with single Active layer can be obtained. The FET_gate layer is used to masking the channel region of conventional FET during the n+ doping. So, the conventional FET type momory cell is fabricated with non-self align method.



(a) (b)

Fig. 3. 1. Mask layout of fabricated (a) conventional FET type and (b) gated-diode memory cells.

Fig. 3. 2 shows the fabrication process flow of the SONOS gated-diode memory cell utilizing the GIDL current.

Fig. 3. 2 (a) : Oxide is deposited on the *p*-type Si-substrate and patterned as Active layer. After that, nitride is deposited for side-wall spacer.

Fig. 3. 2 (b) : Nitride side-wall spacer is formed by dry etching of nitride. The oxide is removed by wet etching with BHF solution. And then, Si fin is formed by inductive coupled plasma (ICP) dry etch process with nitride spacer hard mask. Before etching the Si fin, nitride hard mask at the region of Open layer shown in Fig. 3. 1 is etched to cut off the connection of Si fin.

Fig. 3. 2 (c) : The sides of Si fin is slightly etched by chemical dry etch (CDE) to remove the damaged surface of Si fin and shrink the fin width. The nitride hard mask is removed by wet etching with H_3PO_4 : D.I solution. After that, *p*-type doping is done for preventing leakage current through the surface of substrate. And then, oxide is deposited by High density plasma chemical vapor deposition (HDPCVD) process to fill the gap between Si fins. Here, the thickness of the oxide which is deposited on the narrow Si fin is very small while the gap is filled with oxide due to the characteristics of HDPCVD (deposition-sputtering- redeposition process for gap fill). Thanks to this oxide profile, surface planarization can be easily obtained without chemical mechanical polishing (CMP) process.

Fig. 3. 2 (d) : The oxide is etched by BHF solution to reveal the top surface of Si fin. And then, n^+ diffusion region is formed at the top of Si fin by ion implant process. Before

ion implant process, channel region of conventional FET type momory cell is prevented by PR using FET_gate layer shown in Fig. 3. 1. After ion implant process, annealing process with temperature of 900 °C for 30 minutes and temperature of 1050 °C for 5 seconds is done.

Fig. 3. 2 (e) : The oxide is etched by BHF solution to reveal the p-n^+ junction for overlapping the p-n junction with the gate. Here, the remained oxide prevents the electrical connection between active Si fins. And then, tunneling oxide / trapping nitride / blocking oxide are deposited. Here, the thickness of gate stack is 3nm/6nm/9nm, respectively.

Fig. 3. 2 (f) : Doped poly silicon for gate material is deposited and patterned. The back-end-of-line (BEOL) process follows gate formation; inter-layer dielectric (ILD) deposition, contact formation and metallization.

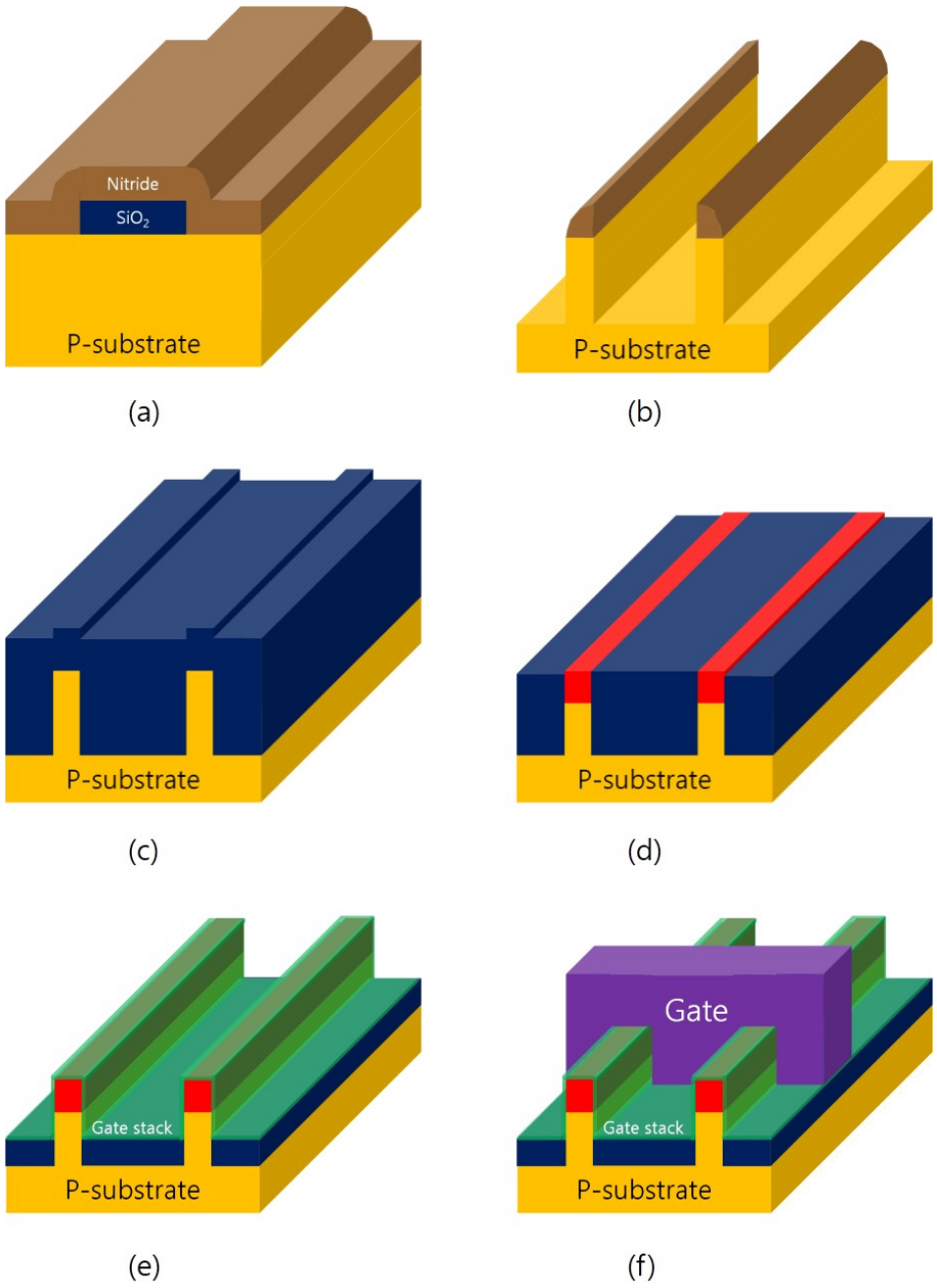


Fig. 3. 2. Fabrication process flow of the SONOS Gated-diode memory cell utilizing the GIDL current.

There are two other methods to form the fin structure.

Fig. 3. 3 shows the fabrication process flow of the SONOS gated-diode memory cell utilizing the GIDL current using the nitride wet etch and CMP process.

Fig. 3. 3 (a) : The nitride and oxide are deposited on the p-type Si substrate. After the Active layer patterning, oxide and nitride are etched by anisotropic dry etching. After that, N+ diffusion region is formed by ion implantation into the wafers being tilted 30°. After ion implant process, annealing process with temperature of 900 °C for 30 minutes and temperature of 1050 °C for 5 seconds is done.

Fig. 3. 3 (b) : The active Si fin is etched by ICP dry etch process with oxide hard mask. The sides of Si fin are slightly etched by CDE process to remove the damaged surface of Si fin. And then, the nitride is etched by wet etching with H_3PO_4 : D.I solution. Here, the oxide on the nitride makes it possible to etch sides of nitride only and the length of the etched nitride will be the fin width.

Fig. 3. 3 (c) : The oxide on the nitride is removed by wet etching process with BHF solution. And then, oxide is deposited by HDPCVD process to fill the gap between Si fins and CMP process is done for surface planarization. Here, the nitride is used as a stopping layer of CMP process. The nitride is removed by wet etching with H_3PO_4 : D.I solution. As show in the figure, the region of etched nitride is filled with oxide.

Fig. 3. 3 (d) : The revealed region of Si fin is etched by ICP dry etch process with oxide hard mask. As shown in the figure, Si fin is devided into the two narrow Si fins with n^+ diffusion region at the top region. After that, *p*-type doping is done for preventing

the electrical connection between two fins.

Fig. 3. 3 (e) : The oxide is etched by BHF solution to reveal the p-n^+ junction for overlapping the p-n^+ junction with the gate. Here, the remained oxide prevents the electrical connection between active Si fins. And then, tunneling oxide / trapping nitride / blocking oxide are deposited. Here, the thickness of gate stack is 3nm/6nm/9nm, respectively.

Fig. 3. 3 (f) : Doped poly silicon for gate material is deposited and patterned. The back-end-of-line (BEOL) process follows gate formation; inter-layer dielectric (ILD) deposition, contact formation and metallization.

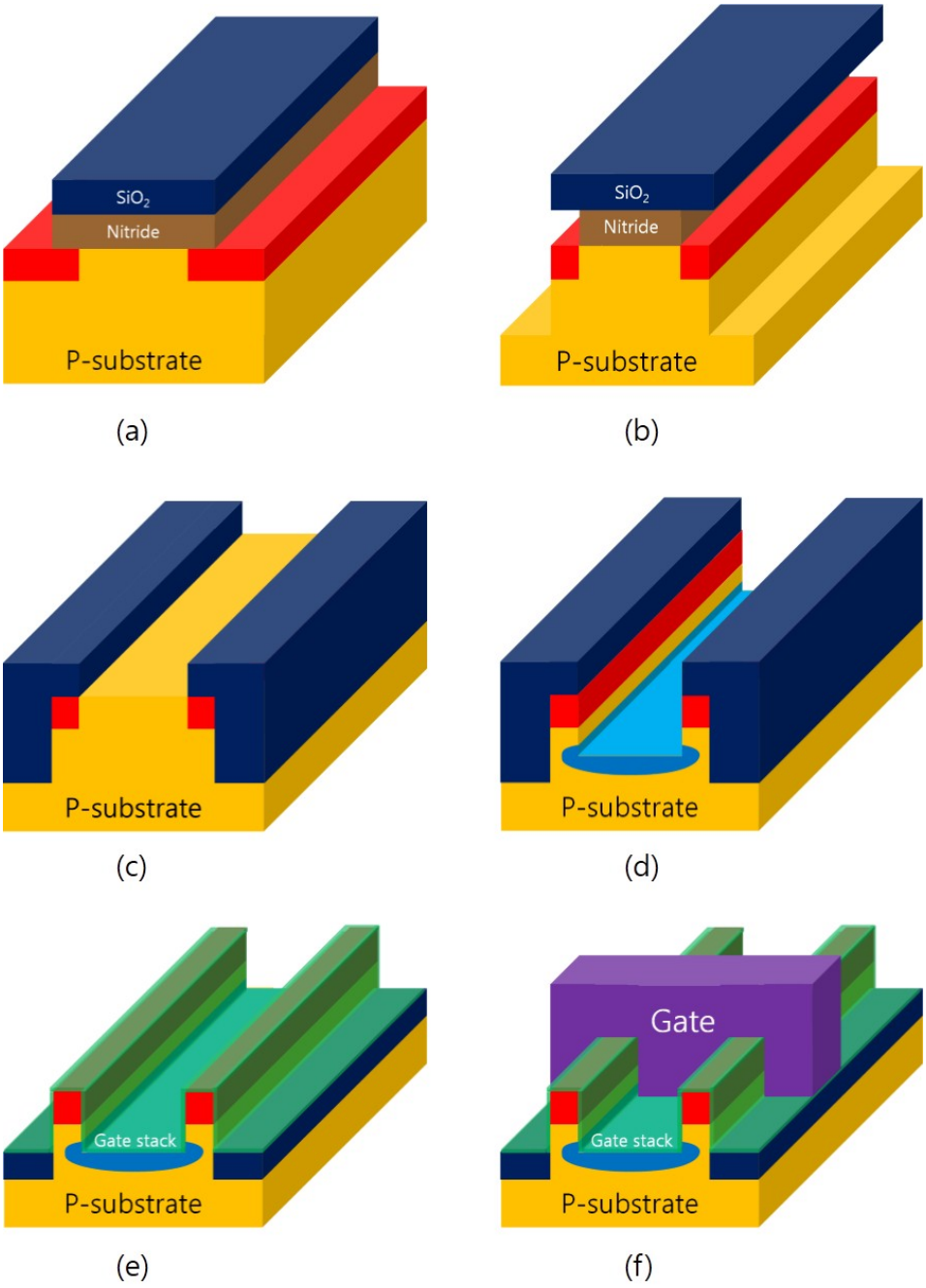


Fig. 3. 3. Fabrication process flow of the SONOS gated-diode memory cell utilizing the GIDL current using the nitride wet etch and CMP process.

Fig. 3. 4 shows the fabrication process flow of the SONOS gated-diode memory cell utilizing the GIDL current using the CMP process and nitride spacer.

Fig. 3. 4 (a) : The nitride is deposited on the *p*-type Si substrate. After the Active layer patterning, nitride and Si are etched by anisotropic dry etching. And the sides of Si fin are slightly etched by CDE process to remove the damaged surface of Si fin.

Fig. 3. 4 (b) : The oxide is deposited by HDPCVD process to fill the gap between Si fins and CMP process is done for the surface planarization.

Fig. 3. 4 (c) : The nitride is removed by wet etching with H_3PO_4 : D.I solution. And then, N+ diffusion region is formed by ion implantation into the revealed Si fin. After ion implant process, annealing process with temperature of 900 °C for 30 minutes and temperature of 1050 °C for 5 seconds is done.

Fig. 3. 4 (d) : The nitride is deposited and etched to make side-wall spacer. Here,

Fig. 3. 4 (e) : The revealed region of Si fin is etched by ICP dry etch process with nitride spacer hard mask. As shown in the figure, Si fin is devided into the two narrow Si fins with n^+ diffusion region at the top region. After that, *p*-type doping is done for preventing the electrical connection between two fins.

Fig. 3. 4 (f) : The nitride and oxide are etched by wet etch process with H_3PO_4 : D.I solution and BHF solution, respectively. And then, tunneling oxide / trapping nitride / blocking oxide are deposited. Doped poly silicon for gate material is deposited and patterned. The back-end-of-line (BEOL) process follows gate formation; inter-layer dielectric (ILD) deposition, contact formation and metallization.

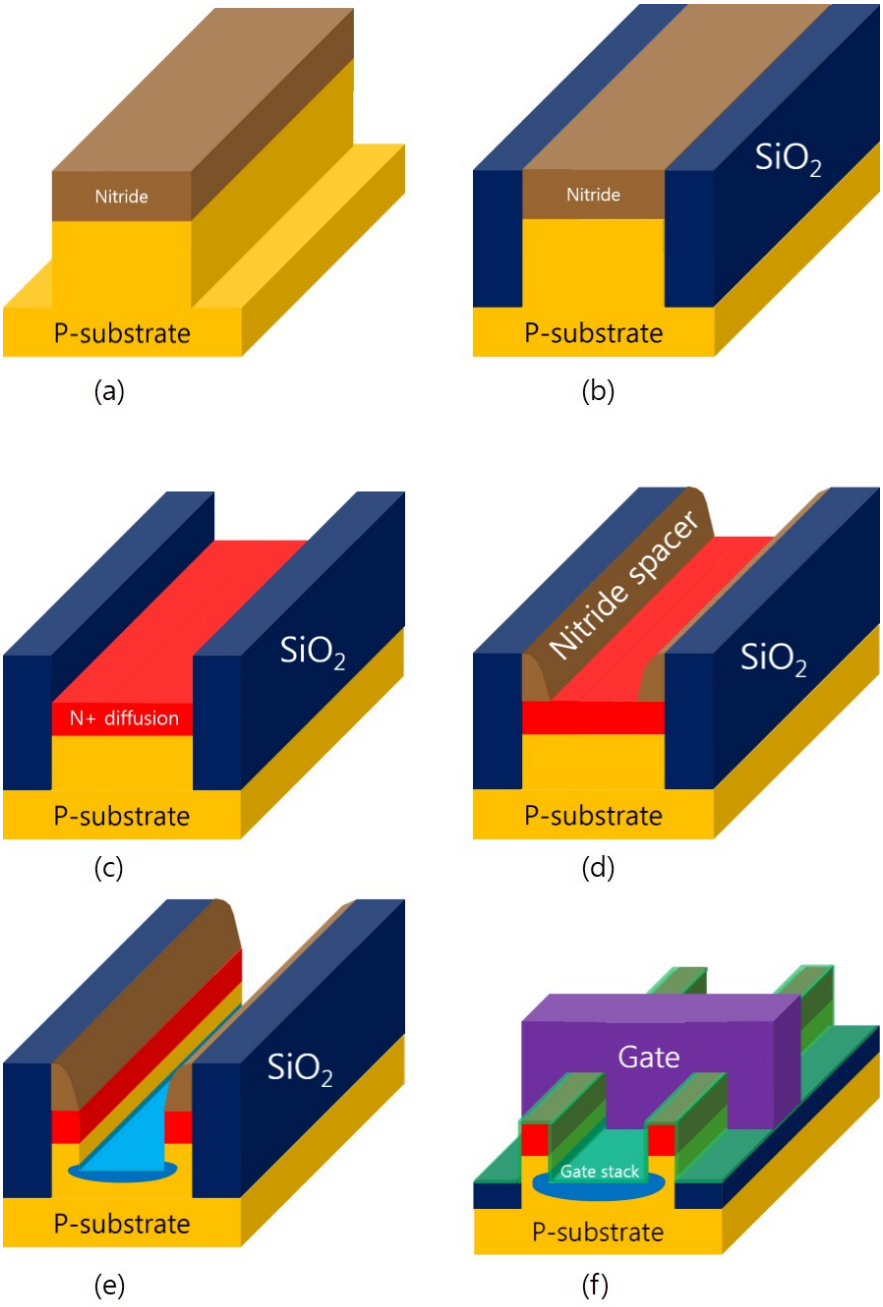


Fig. 3. 4. Fabrication process flow of the SONOS gated-diode memory cell utilizing the GIDL current using the CMP process and nitride spacer.

3.2 SEM Images of Fabricated Structure and Key Process Issues

Firstly, third process method shown in Fig. 3. 4 is used for fabrication. Fig. 3. 5 shows the scanning electron microscope (SEM) image of fabricated twin-fin structure which is corresponding to the structure shown in Fig. 3. 4 (e).

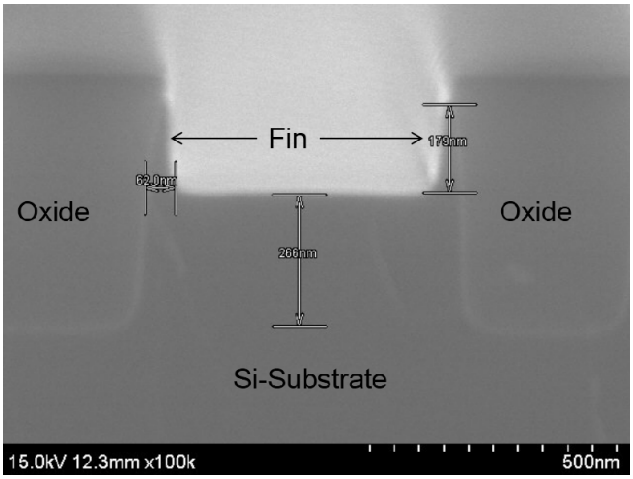


Fig. 3. 5. The scanning electron microscope (SEM) image of fabricated twin-fin structure which is corresponding to the structure shown in Fig. 3. 4 (e).

As shown in Fig. 3. 5, fabricated twin-fin structure is unstable and the variation of W_{fin} in the wafer is severe. This is originated from the bad uniformity of CMP process. After the CMP process, the variation of the thickness of nitride stopping layer in the

wafer is severe. So, the height difference of oxide after nitride strip is also variable. This originates the variation of nitride spacer structure. Therefore, the key process step in the third process method shown in Fig. 3. 4 is uniform CMP process. In order to reduce the critics of uniformity of CMP process, second process method shown in Fig. 3. 3 is devised. In the second process method, W_{fin} is defined as the length of etched nitride by wet etching with H_3PO_4 : D.I solution. Fig. 3. 6 shows the SEM images of fabricated structure which is corresponding to (a) the Fig. 3. 3 (b), (b) the Fig. 3. 3 (c) and (c) the Fig. 3. 3 (e), respectively.

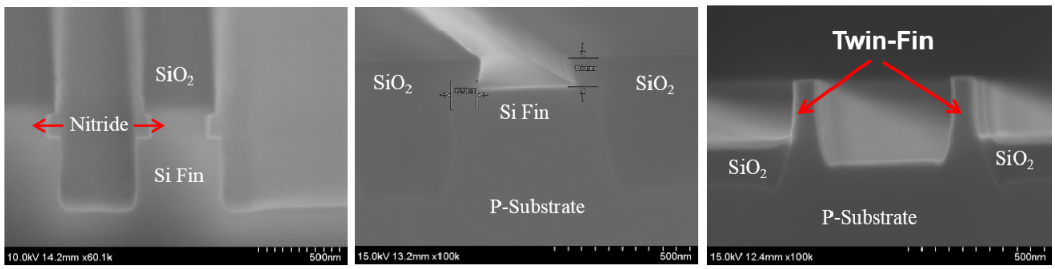


Fig. 3. 6. SEM images of fabricated structure which is corresponding to (a) the Fig. 3. 3 (b), (b) the Fig. 3. 3 (c) and (c) the Fig. 3. 3 (e).

As shown in Fig. 3. 6 (a), the sides of nitride on the Si fins are recessed by the isotropic wet etch process. The etch rate of nitride in the H_3PO_4 : D.I solution is about 3 nm/min. This is controllable for narrow fin width (<100 nm). As shown in Fig. 3. 6 (b), the region of etched nitride is filled with oxide and this oxide plays a role of hard mask of Si etch process for twin fin. The width of oxide hard mask on the Si fin is about 90 nm and the uniformity in the wafer is good. As shown in Fig. 3. 6 (c), twin fin is well made and W_{fin} is also about 90 nm. Here, W_{fin} can be shrunk by CDE process. But there are still some weakness to make narrow fin with n^+ diffusion region. In the second process method, n^+ diffusion region is formed at the beginning of the process steps. So, n^+ diffusion region undergoes the CDE process and wet etching with the H_3PO_4 : D.I solution. These process steps can etch the n^+ diffusion region and the etch speed is faster than un-doped Si region. So, there is a risk that n^+ diffusion region is getting much smaller than expected or disappear. In order to eliminate the risk, the first process method shown in Fig. 3. 2 is devised. The narrow twin fin is formed by nitride spacer at the beginning of the process steps and n^+ diffusion region is formed after CDE process and wet etching with the H_3PO_4 : D.I solution. So, this process method is best suitable for fabrication of gated-diode memory cell and array. Fig. 3. 7 shows the SEM images of fabricated structures using first process method shown in Fig. 3. 2. The structure of Fig. 3. 7 (a) is corresponding to the Fig. 3. 2 (b) and the width of nitride spacer and Si fin is about 90 nm. After the CDE and nitride wet strip processes, the W_{fin} is shrunk to about 50 nm as shown in Fig. 3. 7 (b). And the height of fin is about 300 nm. After the HDPCVD

process, the gap between Si fins are filled with oxide and the thickness of the oxide on the narrow fins is very small as shown in the Fig. 3. 7 (c). Thanks to the oxide profile, the surface planarization can be easily obtained by only wet etch process with BHF solution without CMP process as shown in Fig. 3. 7 (d). At the same time, the top of Si fins is revealed. So, n^+ diffusion region at the top of Si fins can be formed stably by ion implantation process. After forming n^+ diffusion region, only oxide wet etch process is done and gate stack is deposited as shown in Fig. 3. 7 (e).

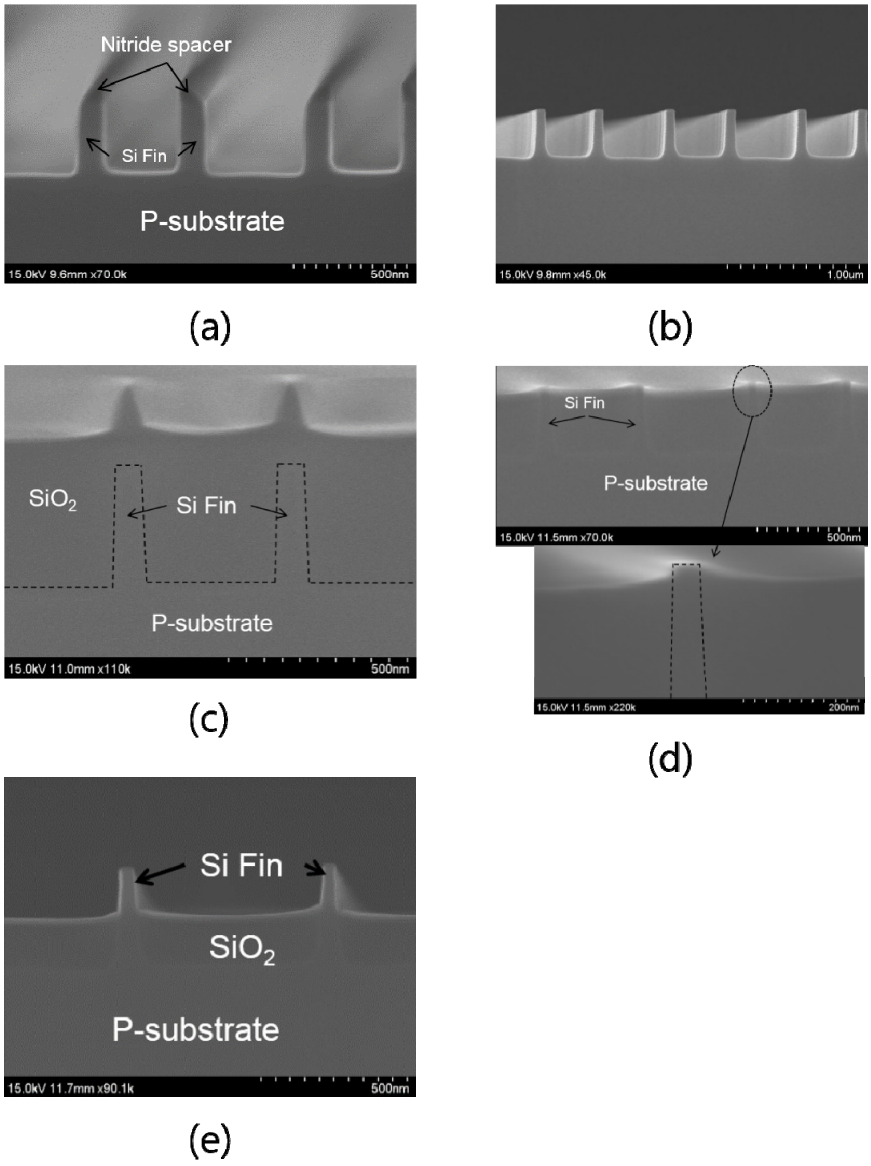
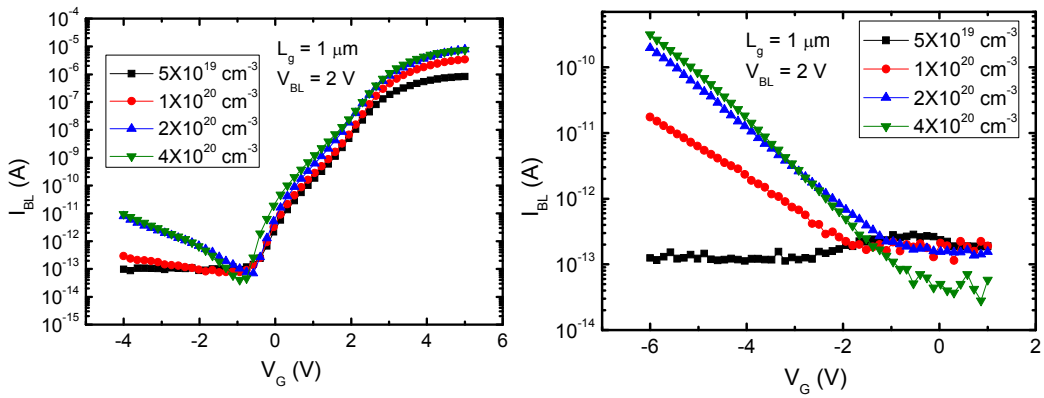


Fig. 3. 7. SEM images of fabricated structure using first process method shown in Fig. 3. 2.

3.3 Electrical Characteristics of SONOS Gated-diode Memory Cell and Array utilizing GIDL Current

Electrical characteristics of fabricated SONOS gated-diode memory cell and array utilizing GIDL current are discussed in this section. As mentioned in the previous chapter, fabricated devices are implemented with twin fin structure. Electrical characteristics are analized by using B1500 and HP4156C parameter analyzer.

First, the electrical characteristics with variations of peak doping concentration of n^+ diffusion region are investigated in conventional SONOS FET type memory and gated-diode memory cell.

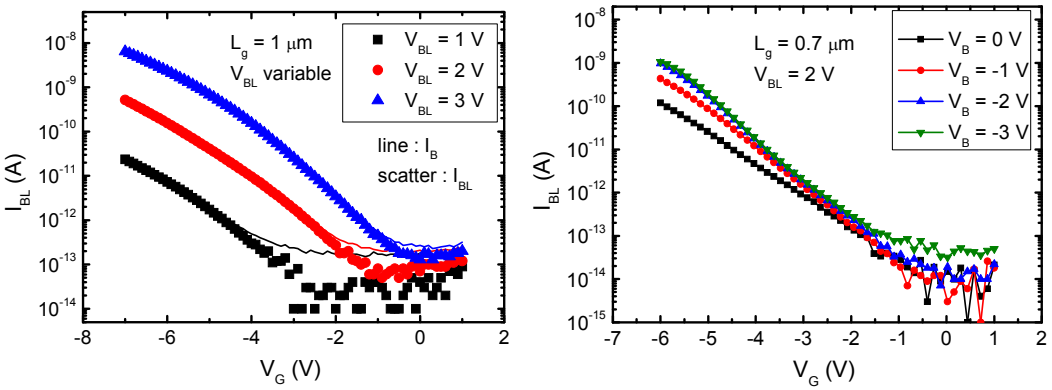


(a) (b)

Fig. 3. 8. Measured I-V characteristics with variations of peak doping concentration of n^+ diffusion region in the (a) conventional FET type and (b) gated-diode memory cell

As shown in Fig. 3. 8, the channel current increases with increased doping concentration. And the GIDL current in both conventional FET type and gated-diode memory cell also increases as the doping concentration increases. When the doping concentration is low as 5×10^{19} cm^{-3}, the GIDL current is not generated due to the wide depletion width. When the doping concentration is 10^{20} cm^{-3}, the measured GIDL current shows the variation through the whole wafer. When the doping concentration is 2×10^{20} cm^{-3} and 4×10^{20} cm^{-3}, the measured GIDL current is reliable and similar.

To investigte the electrical characteristics of fabricated SONOS gated-diode memory cell using GIDL current, the bit-line current (I_{BL}) and body current (I_B) are measured with variable bit-line voltage (V_{BL}) as shown in Fig. 3. 9 (a). As V_{BL} increases, I_{BL} increases. And the measured I_{BL} and I_B are the same. But the leakage current of I_B is larger than that of I_{BL} beacause I_B flows the back-side of whole wafer.



(a) (b)

Fig. 3. 9. The electrical characteristics of fabricated SONOS gated-diode memory cell with variations of (a) V_{BL} and (b) V_B.

As V_B decreases, I_{BL} increases as shown in Fig. 3. 9 (b). This is caused by increased lateral electric field (E-field) between bit-line and body. As known before, the GIDL current is affected by lateral E-field as well as vertical E-field.

To investigate the programming characteristics, the incremental step pulse programming (ISPP) is performed as shown in Fig. 3. 10. For FN tunneling used as programming mechanism, the gate voltage of program pulse V_{pp} is increased by a constant value after each program step. As the V_{pp} increases, the trapped electrons at the trapping nitide increases, and then, I_{BL} also increases.

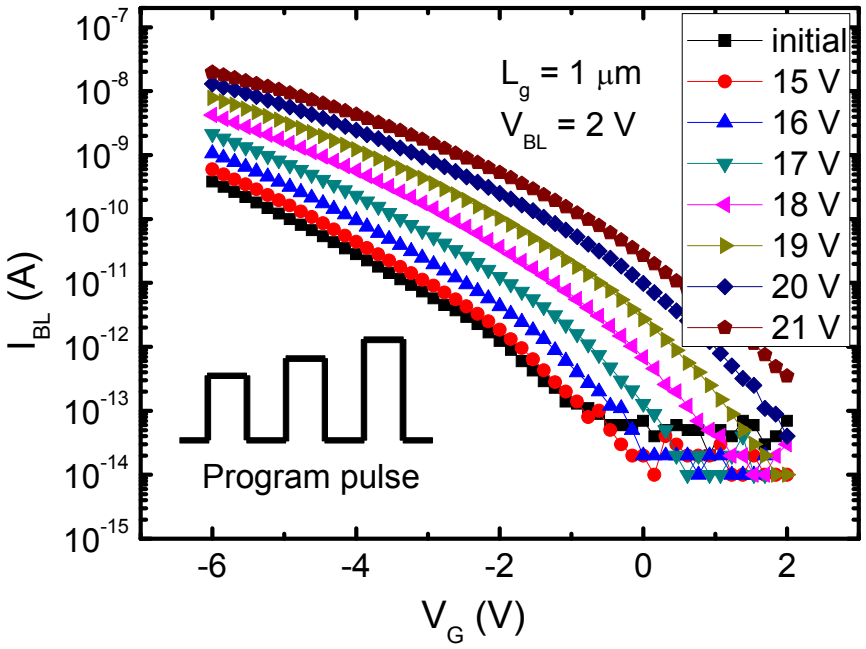


Fig. 3. 10. ISPP operation in the fabricated SONOS gated-diode memory cell.

To perform the erase from the programmed state, the negative gate voltage pulse is applied with a variation of V_{BL} and the results are shown in Fig. 3. 11. After the device is programmed with V_{pp} of 20V, the erase pulse voltage of -9 V is applied with a variation of V_{BL}. As shown in the results, the erase is not performed when the V_{BL} is smaller than 3 V. This indicates that holes are not supplied from the p-type body. So, we can guess that the p-n junction is much depper than expected due to the thermal process. As the V_{BL} is getting high, generated holes with GIDL current enhance the erase performance.

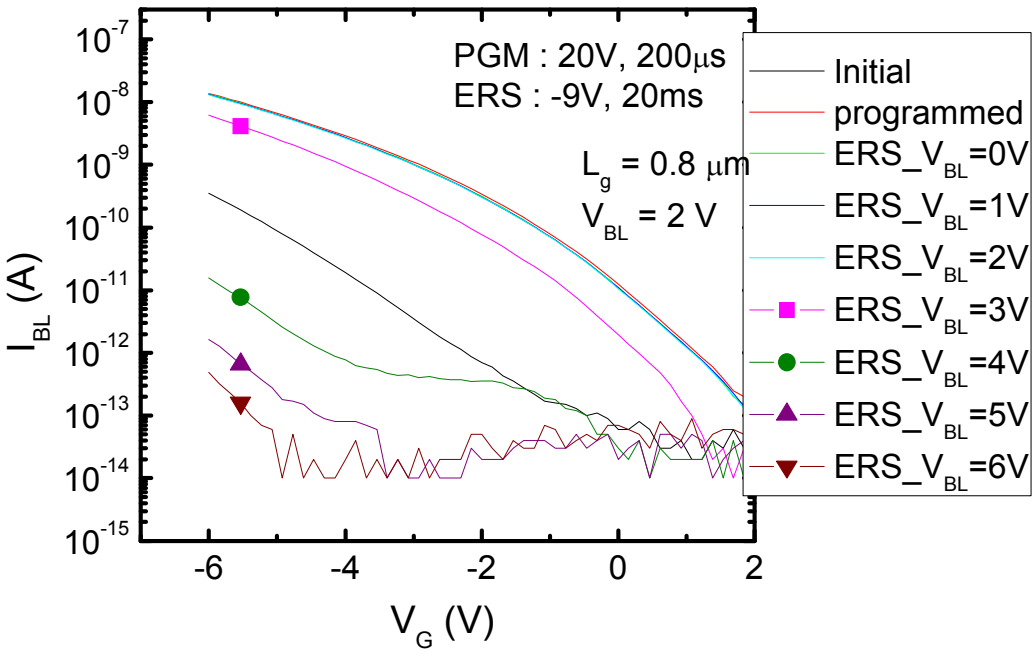
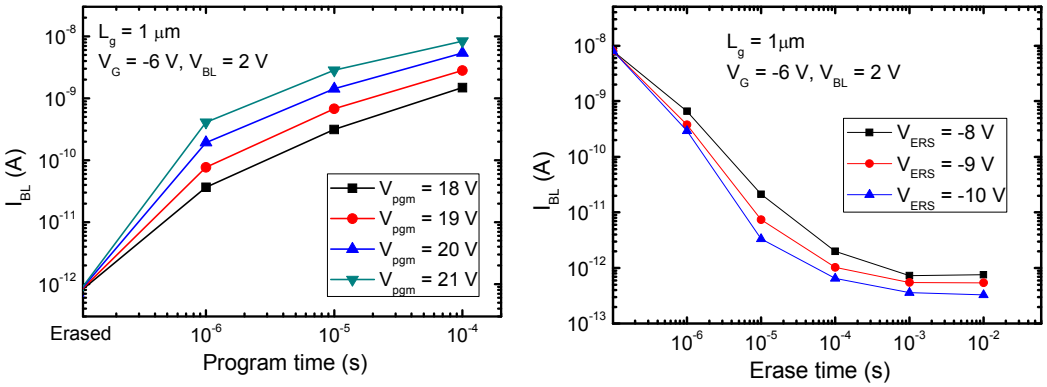
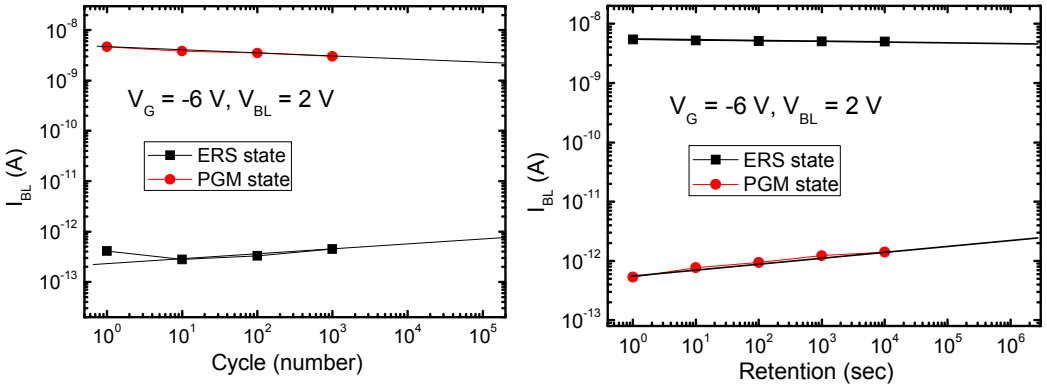


Fig. 3. 11. Erase characteristics with a variation of V_{BL} during the erase pulse is applied.



(a) (b)

Fig. 3. 12. (a) Program and (b) erase characteristics with a variation of pulse time and voltage.



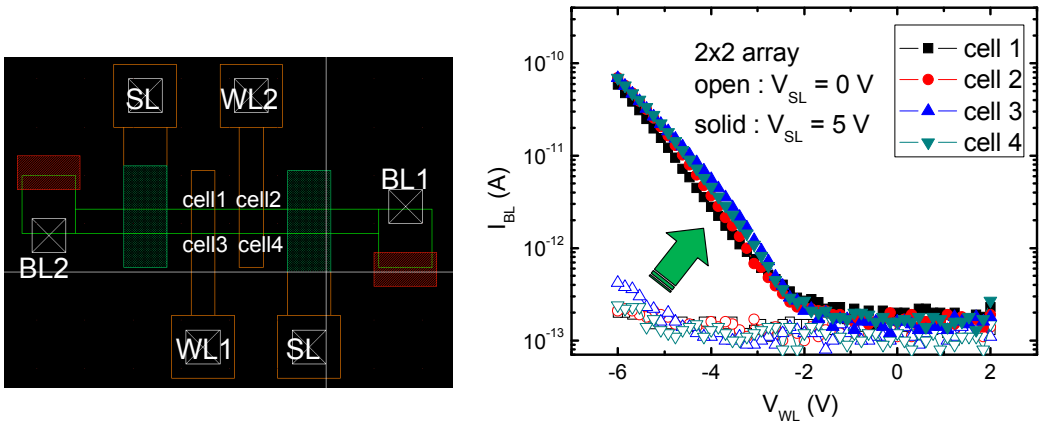
(a) (b)

Fig. 3. 13. (a) Cycling and (b) retension characteristics of fabricated device.

Fig. 3. 12 (a) and (b) show the program and erase characteristics with a variation of pulse time and voltage. The reasonable program/erase speed and pulse voltage are shown.

Fig. 3. 13 (a) and (b) show the cycling and retension characteristics of fabricated device. The reliable characteristics are shown.

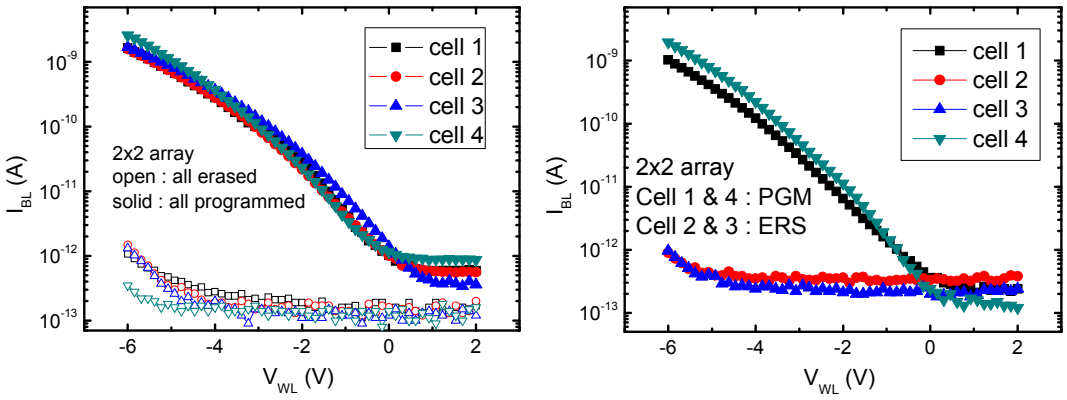
Finally, the electrical characteristic of fabricated aray is investigated. The array of 2X2 is fabricated with single active layer, two word-lines (WL) and two select transistors (SL) as shown in Fig. 3. 14 (a). First, cell currents of all cells are measured with a variation of voltage of select transistors. As shown in Fig. 3. 14 (b), cell currents are controlled by select transistors well. And measured cell currents of each cell are quite similar.



(a) (b)

Fig. 3. 14. (a) Lay out of 2X2 array (b) operation of select transistor in array.

The program/erase characteristic in 2×2 cell aray is investigated in Fig. 3. 15. The cell currents of all cells which have programmed state are similar and cell currents of all cells which have erased state are also similar. To verify the selective program/erase characteristic in 2×2 cell arays, cell currents are measured after cells 1 and 4 are programmed and cells 2 and 3 are erased. As shown in Fig. 3. 15 (b), the cells can be selectively programmed or erased well.



(a) (b)

Fig. 3. 15. Program/erase characteristics of 2×2 cell array. (a) programming or erasing of all cells and (b) selective programming or erasing in array.

Chapter 4

Optimization of Gated-Diode Memory

In this chapter, several methods and analysis are demonstrated to increase the GIDL current for large enough read cell current in gated-diode memory through a device simulation. First, comparison between measurement and simulation is performed to secure a confidence of simulation results. And variable structures with SiGe which has low band gap value are investigated to increase the band-to-band generation rate.

4.1 Comparison between Measurement and Simulation

Fig. 4. 1 shows the device structure which is used in the device simulation. Here, the Sentaurus TCAD device simulation tool is used. The simulated structure has fin width of 50 nm, gate length of 500 nm, peak doping concentration of n^+ diffusion region of 4×10^{20} cm^{-3}, junction depth of 70 nm, and n^+ poly silicon gate. The gate stack of ONO have the thickness of 3nm, 6nm and 9nm, respectively.

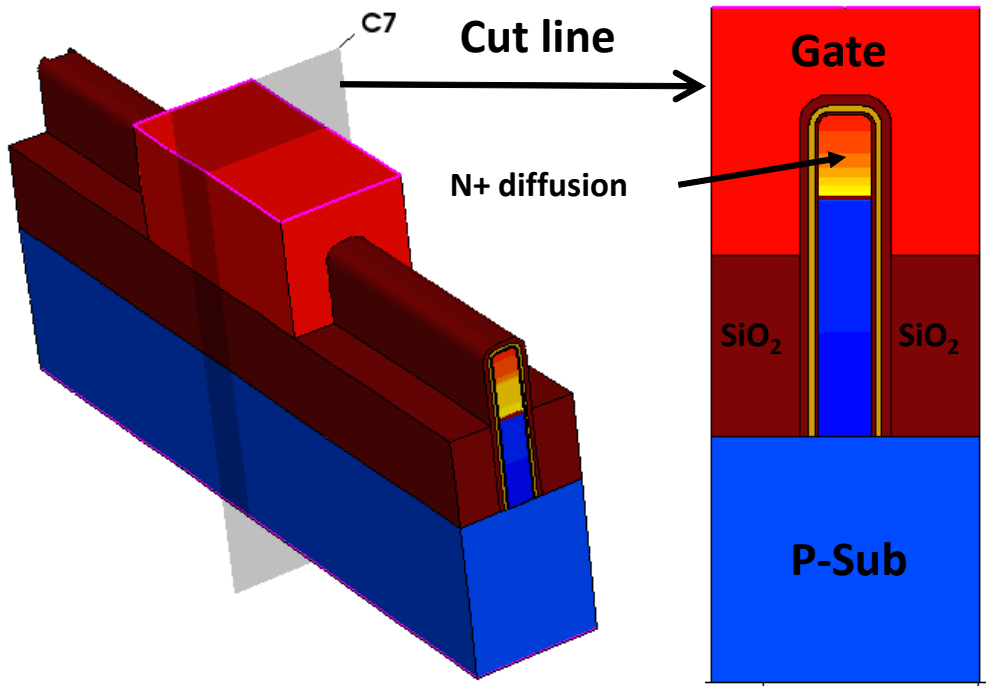
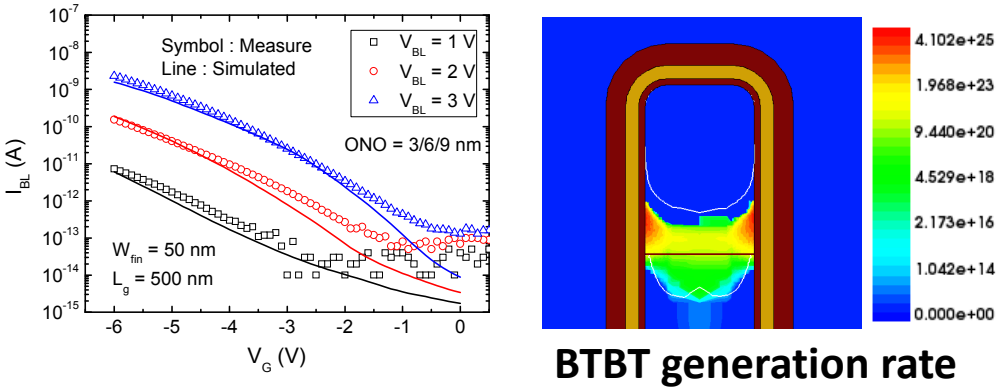


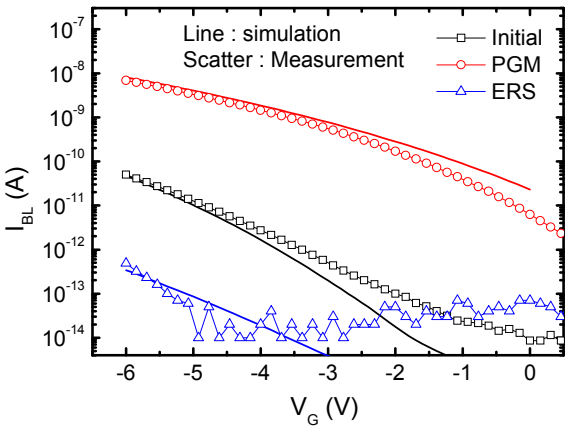
Fig. 4. 1. Device structure which is used in the simulation.



(a) (b)

Fig. 4. 2. (a) Comparison of simulated I_{BL}-V_G curves with measured one. (b) cut-plane which shows the BTBT generation rate at bias condition of V_{BL} = 2 V and V_G = -6 V.

Firstly, the simulated I_{BL}-V_G curves with a variation of V_{BL} are compared with measured ones. As shown in Fig. 4. 2 (a), simulated data is similar with measured one especially at the high $|V_G|$. At low $|V_G|$, the trap assisted tunneling will be dominant than BTBT, then the difference between simulation and measurement occurs. Fig. 4. 2 (b) shows the cut-plane of the top in the fin along the gate direction and the BTBT generation rate is represented. As shown in the figure, BTBT generation occurs only at the surface adjacent to the p-n^+ junction while the BTBT generation does not occur at the edge and top of the fin due to the high doping concentration.



	Measurement	Simulation
PGM	V_{pp} = 20 V V_{BL} = 0 V t_{PGM} = 100 μs	Fixed charge of -2x10^{19}/cm^3 in nitride layer
ERS	V_{EP} = -9 V V_{BL} = 6 V t_{ERS} = 20 ms	Fixed charge of 2x10^{19}/cm^3 in nitride layer

Fig. 4. 3. Comparison of I_{BL}-V_G curves between measurement and simulation when the device states are initial, programmed, and erased.

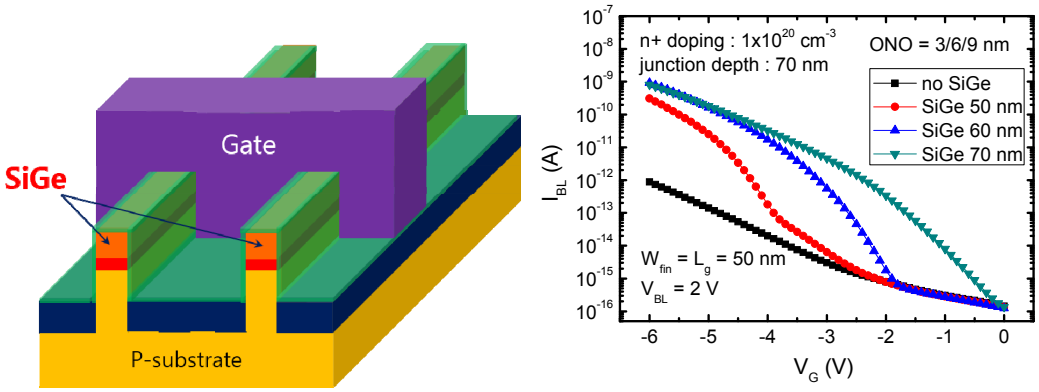
Fig. 4. 3 shows the comparison of I_{BL}-V_G curves between measurement and simulation when the device states are initial, programmed, and erased. And the conditions of PGM/ERS for measurement and fixed charge in nitride trapping layer for simulation are shown in the table. From the results, the stored charge density at the nitride trapping layer can be estimated as about 2×10^{19} cm^{-3}.

4.2 Increase of GIDL Current using SiGe

In order to increase the GIDL current, SiGe which has lower bandgap value than Si is used. The lattice constant of the Si is around 5.530 Å while for Ge is around 5.658 Å which shows that Ge atoms are bigger than the Si atoms (4.19%). So the bandgap of Ge is smaller (0.66 eV for Ge and 1.12 eV for Si).). The mixture of these atoms, which form an alloy as $Si_{1-x}Ge_x$, is suitable for sacrificial layer.

After SiGe epitaxial growth on the p-type substrate, device fabrication process introduced in the chapter 3 follows. And then, the device structure as shown in Fig. 4. 4 (a) can be achieved. Here, SiGe is also doped as n^+ diffusion layer and 30% of Ge fraction (x=0.3) is used for the simulation. And W_{fin} and L_g are 50 nm in order to verify the GIDL current level when the device size is decreased. As shown in the simulation result of Fig. 4. 2, BTBT generation occurs adjacent to p-n junction. So, the relation between thickness of SiGe and depth of p-n^+ junction is key parameter. Then, simulation is performed with p-n junction depth of 70 nm and a variation of SiGe thickness. By using the SiGe layer, the GIDL current is greatly increased as shown in Fig. 4. 4 (b). But the *I-V* characteristics varies with a variation of the thickness of SiGe. From the result, we can know that the thickness of SiGe should be larger than the depth of p-n^+ junction for the generation of BTBT at the SiGe region. Therefore, the junction control is very critical in this structure. And there is another weakness in the fabrication process of this structure.

The damages on the SiGe layer during the process like as CDE and SC1 cleaning is critical because the etch rate of SiGe in those processes is much larger than Si. Moreover, BTBT generation still does not occur at the region of the edge and the top of fin nevertheless the SiGe layer is used on the top of fin.

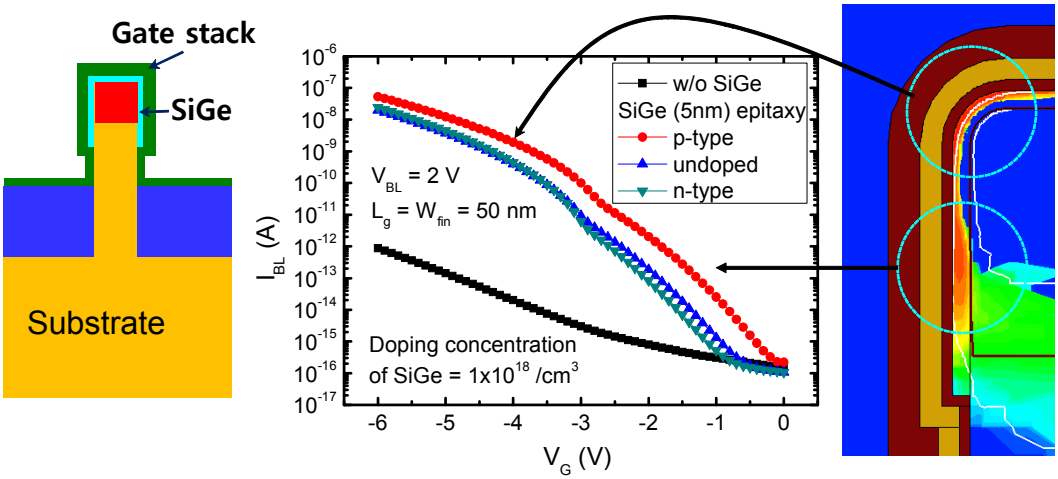


(a) (b)

Fig. 4. 4. (a) Device structure with SiGe layer on the top of fin (b) simulation results with a variation of thickness of SiGe.

The other method to use the SiGe layer is devised for elimination of these weakness. The SiGe is formed by selective epitaxial growth (SEG) after forming of n^+ diffusion

region in the fabrication process shown in Fig. 3. 2. So, the critical process steps are done before the SiGe layer is formed. Also, process is so simple that only SiGe SEG process is added ahead of gate stack deposition. The cut-plane of the structure after SiGe SEG and gate stack deposition is shown in Fig. 4. 5 (a). Fig. 4. 5 (b) shows the simulated results with SiGe SEG layer. Here, the thickness of SiGe layer is 5 nm and doping type of SiGe is varied as *p*-type, un-doped, and *n*-type.



(a) (b) (c)

Fig. 4. 5. (a) Cut-plane of device structure with SiGe SEG process (b) simulation results with SiGe SEG layer with a variation of doping in SiGe (c) BTBT generation rate at the bias condition of V_G = -6 V and V_{BL} = 2 V.

As shown in the results, the GIDL current is greatly increased thanks to the SiGe layer. In addition, BTBT generation occurs at the edge of the fin as shown in Fig. 4. 5 (c)

which shows the BTBT generation rate in the structure with p-type SiGe layer at a bias condition of V_G = -6 V and V_{BL}= 2 V. Also, current hump is shown due to the corner effect at the edge of the fin. The *p*-type SiGe layer shows largest GIDL current and un-doped and *n*-type SiGe layer show similar GIDL current. This results can be understood with the relation of surface potential and band gap. The energy band diagram at the cut-line of the edge of the fin is shown in Fig. 4. 6. The E_g and $|\psi_s|$ of p-type, un-doped, and *n*-type SiGe layers are also represented as a table. The BTBT generation rate increases, as the E_g decreases and $|\psi_s|$ increases. The E_g of doped SiGe is a little smaller than that of un-doped SiGe due to the band gap narrowing effect. And the $|\psi_s|$ of *p*-type SiGe is largest and $|\psi_s|$ of un-doped SiGe is larger than that of n-type SiGe. So, the BTBT generation rate with *p*-type SiGe is largest. And the BTBT generation rate with n-type and un-doped SiGe are similar due to the compensation between E_g and $|\psi_s|$.

Here, the $|\psi_s|$ is not much different with different doping type because SiGe layer is fully depleted due to the low doping concentration and narrow thickness even at bias condition of V_G = V_D = 0 V as shown in Fig. 4. 7 (a). As the doping concentration of *p*-type SiGe increases, the $|\psi_s|$ increases drastically as shown in Fig. 4. 7 (b). When the doping concentration is 5×10^{19} cm^{-3}, $|\psi_s|$ is larger than E_g of SiGe even at a bias condition of V_G = V_{BL} = 0 V. So it can increase the off-current. Therefore the optimization of the doping concentration and thickness of SiGe layer is needed.

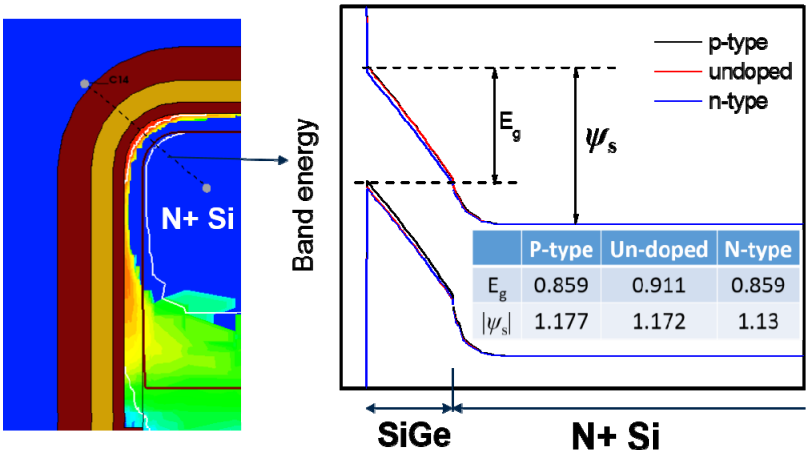
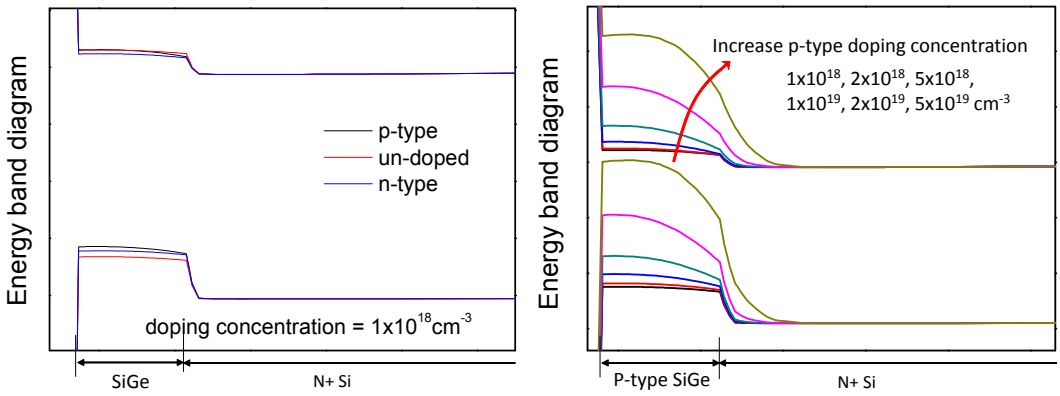


Fig. 4. 6. The energy band diagram at the cut-line of the edge of the fin with a variation of doping type of SiGe layer at a bias condition of V_G = -6 V and V_{BL} = 2 V.



(a) (b)

Fig. 4. 7. The energy band diagram at the cut-line of the edge of the fin (a) with a variation of doping type of SiGe layer and (b) with a variation of *p*-type doping concentration at a bias condition of $V_G = V_{BL}$ = 0 V.

In order to verify the corner effect shown in Fig. 4. 5 (b), the I_{BL}-V_G curves with a variation of radius (r) of the rounded edge of the fin is investigated as shown in Fig. 4. 8. As the r decreases, the current hump is generated at lower $|V_G|$ because the field concentration effect is enhanced at the structure with smaller r. So, the fin structure with smaller r or rectangle shape of edge will help to decrease the operation voltage.

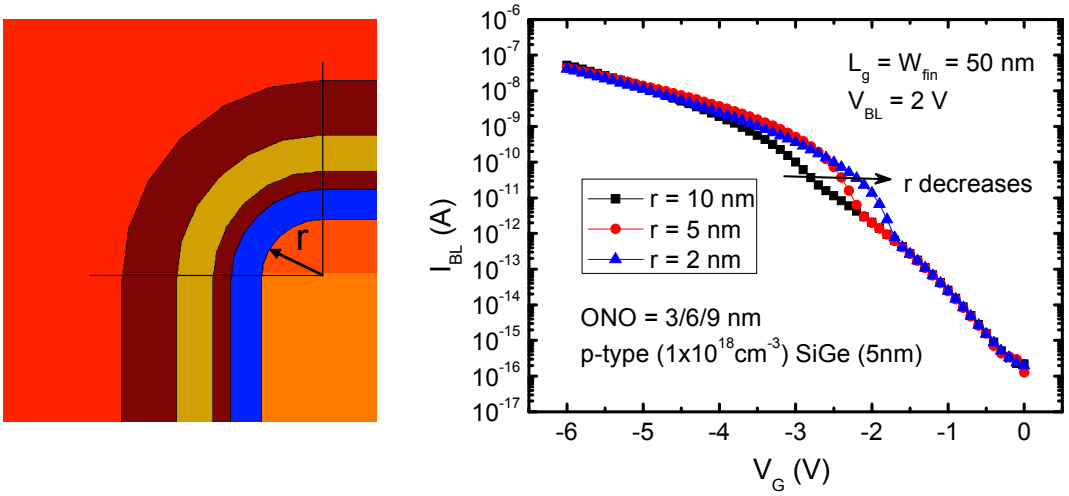
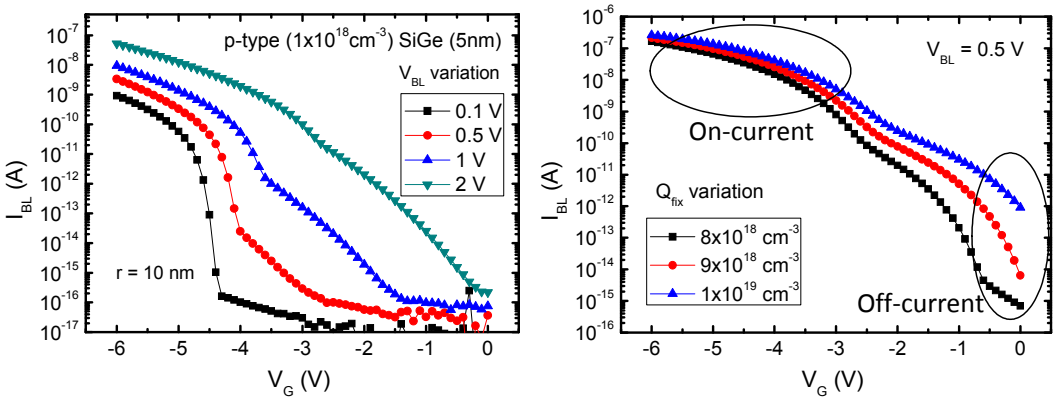


Fig. 4. 8. I_{BL}-V_G curves with a variation of radius (r) of the rounded edge of the fin.

Fig. 4. 9 (a) shows the I_{BL}-V_G curves with a variation of V_{BL}. As the V_{BL} decreases, the corner effect is shown up and the slope is also getting sharper, but the GIDL current decreases. And Fig. 4. 9 (b) shows the I_{BL}-V_G curves with a variation of fixed charge concentration (Q_{fix}) in the nitride trapping layer at a V_{BL} of 0.5 V. As the Q_{fix} increases, on-current which is generated at the corner region is slightly increases but the off-current

which is generated at the sides of the fin increases largely. So, the GIDL current at the sides of the fin should be suppressed in order to decrease the off-current.



(a) (b)

Fig. 4. 9. The I_{BL}-V_G curves with a variation of (a) V_{BL} and (b) fixed charge concentration.

Actually, the concentration of injected charge to the trapping layer in the corner region is much larger than that in the sides of fin because of the field concentration effect on the corner region. As shown in Fig. 4. 10, the injected electron concentration at the corner region is much larger than that at the side region. Thanks to the corner effect, the V_{pp} can be reduced.

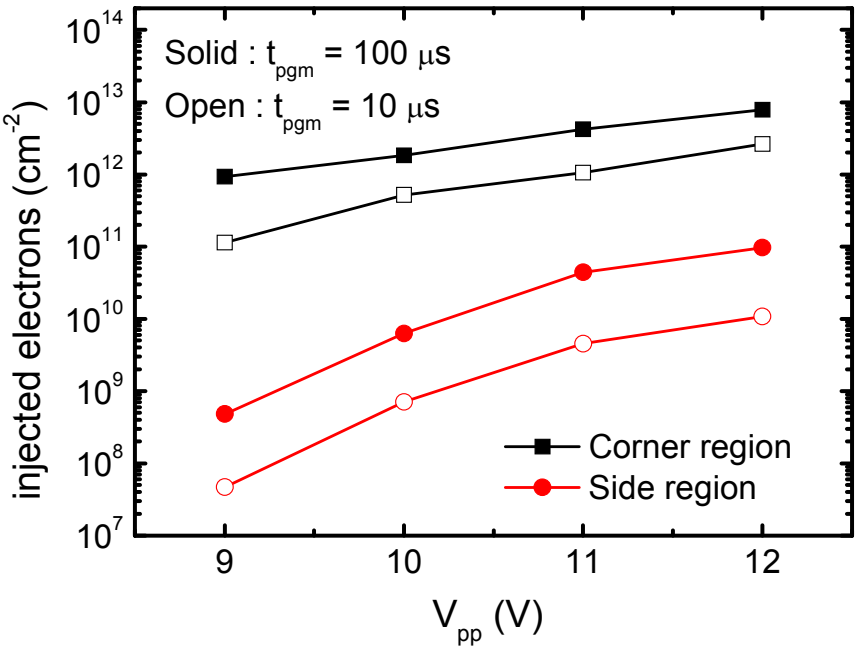


Fig. 4. 10. Injected electrons with a variation of V_{pp} at corner and side region of the fin.

Chapter 5

Conclusions

In this dissertation, gated-diode memory cell and array utilizing gate-induced drain leakage (GIDL) current is proposed, fabricated, and characterized. As the cell size decreases, the short-channel effect (SCE) is the critical issue in conventional flash memory devices. But, there is no SCE in the proposed gated-diode memory. Moreover, the random access can be possible in array structure although the cells are connected serially like as conventional NAND flash memory array. So, the random access and high-density properties can be achieved at the same time.

The gated-diode memory cell and array with SONOS and fin structure are fabricated and characterized. The stable and simple fabrication process for proposed gated-diode memory is also demonstrated. From the measurement of fabricated memory cell and array, stable and reliable memory operation is verified. In addition, gated-diode memory structure with SiGe material is proposed to increase the sensing current and characterized with TCAD simulation.

Appendix

Low Frequency Noise of GIDL Current in MOSFETs

In this chapter, low frequency noise (LFN) including 1/f noise and random telegraph noise (RTN) in the GIDL current is investigated in the MOSFETs. First, the LFN characteristics of GIDL current are compared with those of channel current. And the LFN characteristics of GIDL current are compared between n-type and p-type MOSFETs. Finally, the model which elucidates the generation of RTN in the GIDL current is established.

A.1 LFN of GIDL and Channel Currents

Low frequency noise (LFN) including 1/f noise and random telegraph noise (RTN) due to the capture and emission of single charge at a gate dielectric trap in MOSFETs has been studied [26]-[29]. Until now, most studies on the LFN have been focused on the

channel current in MOSFETs. The threshold voltage fluctuation due to RTN in channel current could be a big issue at nano-scale flash memories [30], and 22 nm SRAM [31]. Nowadays, understanding on the gate induced drain leakage (GIDL) current becomes important in DRAM cells [32], [33], flash memory cells [34], and tunneling FETs [35]. The GIDL in DRAM cells is a key leakage [32] and the RTS-like fluctuation in the GIDL current of Saddle-Fin type DRAM cells also has been investigated to understand cell leakage due to the GIDL [33]. Non-volatile flash memory cells utilizing GIDL has been reported [34]. Recently tunneling FETs have been reported to achieve sub-60-mV/dec by using the band-to-band-tunneling process which is a physical origin of GIDL [35]. Now it is important to understand the GIDL well, and characterizing the RTN in GIDL current is a good approach in investigating the GIDL currents.

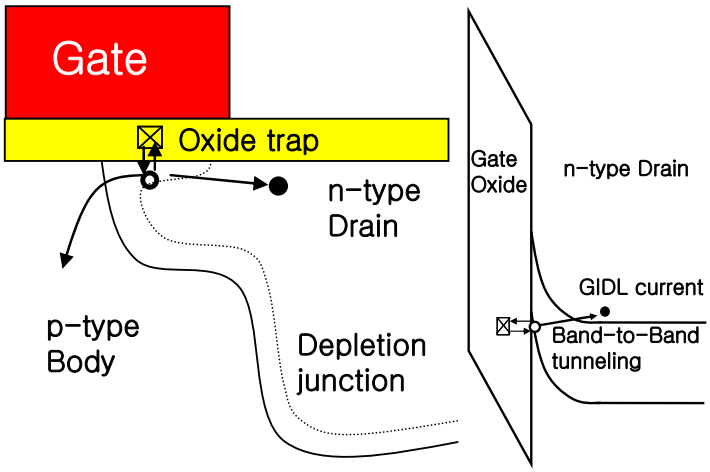


Fig. A. 1. Schematic view and energy band diagram showing trapping and de-trapping of a hole at an oxide trap during GIDL generation in nMOSFETs.

Fig. A. 1 shows the schematic view and energy band diagram showing trapping and de-trapping of a hole at an oxide trap during GIDL generation in nMOSFETs. During the GIDL process, holes generated by the band-to-band tunneling can be trapped into an oxide trap which exists at overlapped drain to gate region. A trapped hole can vary the vertical electric field between gate and drain. So, the trapping/de-trapping of hole can cause RTN in drain and body currents simultaneously. We experimentally observed the same current fluctuation in body and drain currents at a GIDL bias condition and measured I_G, I_D and I_B at the GIDL bias condition are plotted in Fig. A. 2. Here, ΔI_G, ΔI_D, and ΔI_B represent current fluctuation due to noise in I_G, I_D, and I_B, respectively. I_D and ΔI_D are nearly the same with I_B and ΔI_B, respectively. But, we could not observe RTN in I_G, and ΔI_G in I_G are much smaller than those of I_D and I_B.

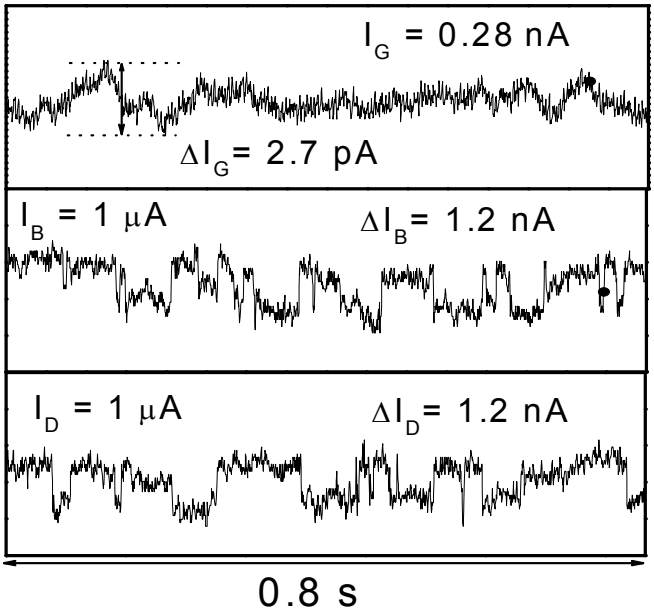
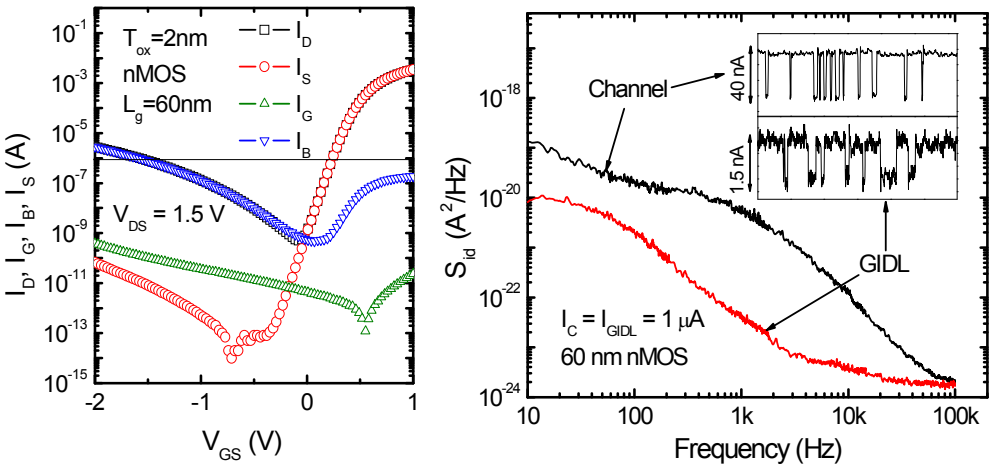


Fig. A. 2. Measured I_G, I_D and I_B at the GIDL bias condition.

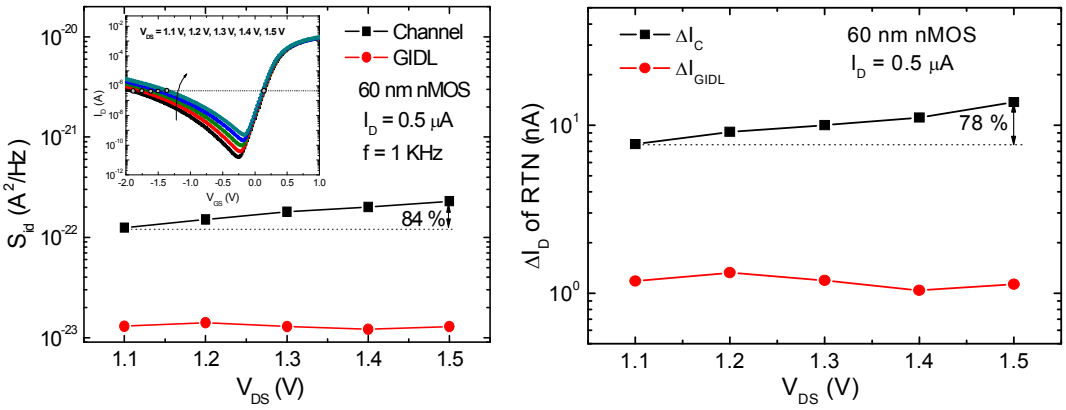


(a) (b)

Fig. A. 3. (a) I-V characteristics and (b) power spectral density (S_{id}) and corresponding RTN waveform in channel (I_C) and GIDL (I_{GIDL}) currents.

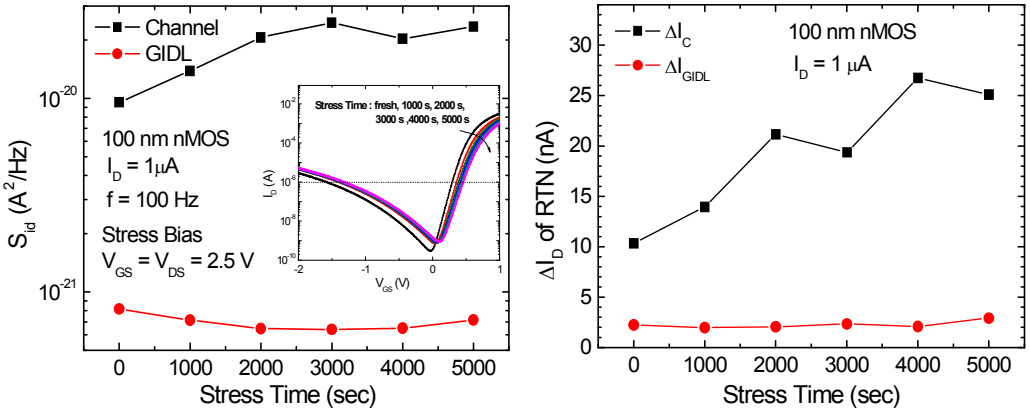
Fig. A. 3 shows the measured (a) I-V characteristics and (b) power spectral density (S_{id}) and corresponding RTN waveform in channel (I_C) and GIDL (I_{GIDL}) currents. Here, I_C is a channel current under sub-threshold bias condition and I_{GIDL} is a GIDL current under GIDL bias condition measured in different devices having the same L_g of 60 nm. Under bias condition for GIDL measurement, the channel of device is in accumulation mode, which means no channel current. So, body current (I_B) and its fluctuation (ΔI_B) are nearly the same as drain current (I_D) and its fluctuation (ΔI_D), respectively. Furthermore, we could observe synchronized RTN waveforms in I_D and I_B at the same time. The gate current at the same bias is much smaller by more than 10^2 times than body current and can be ignored. It is well known that GIDL current is a function of vertical electric field

between the gate and drain. During the GIDL process, electron-hole pairs are generated in the drain overlapped by the gate region, and generated holes move to the body along the surface of drain. A hole can be captured to a trap inside gate dielectric which exists on the drain overlapped with the gate region. A trap with a captured hole gives an effect of more positively charged state and corresponds to lower current level regardless of polarity of the trap. A trap with a captured hole decreases the vertical electric field in a localized surface area and decreases GIDL as a result, which gives low current level in RTN waveform shown in the lower inset of Fig. A. 3 (b). The upper and lower insets stand for channel current fluctuations (ΔI_C) and GIDL current fluctuations (ΔI_{GIDL}), respectively, corresponding to S_{id}s at a current of 1 μA. ΔI_C and ΔI_{GIDL} is 40 nA and 1.5 nA, respectively. The fluctuation due to the RTN in GIDL current is much smaller than that in channel current.



(a) (b)

Fig. A. 4. LFN characteristics with V_{DS} of I_C and I_{GIDL} (a) 1/f noise PSD at 1 kHz. (b) ΔI_D at I_D=0.5 μA.



(a) (b)

Fig. A. 5. LFN characteristics with hot-carrier stress time of I_C and I_{GIDL} (a) 1/f noise PSD at 100 Hz. (b) ΔI_D at I_D=1 μA.

Fig. A. 4 shows LFN characteristics of I_{GIDL} and I_C with a variation of V_{DS} at a gate length (L_g) of 60 nm, when current is the same by controlling V_{GS}. Fig. A. 4 (a) shows the 1/f noise S_{id} and (b) shows the ΔI_D. Here, the S_{id} and ΔI_D of I_{GIDL} are much smaller than those of I_C at each V_{DS}. The S_{id} and ΔI_D of I_{GIDL} are independent on V_{DS} while those of I_C are significantly increasing with increasing V_{DS}. The increase of LFN in I_C with increasing V_{DS} is mainly due to the decrease of effective channel length. The S_{id} and ΔI_D with a variation of channel hot-carrier stress (CHCS) time at an L_g of 100 nm are shown in Fig. A. 4 (a) and (b), respectively. The S_{id} and ΔI_D of I_C are generally increasing with stress time while those of I_{GIDL} are independent of stress time. This indicates that the LFN of I_{GIDL} is not affected by CHCS. Compared to the LFN of channel current, the LFN of GIDL current was much smaller at the same current and showed less drain voltage

dependency and better stress immunity.

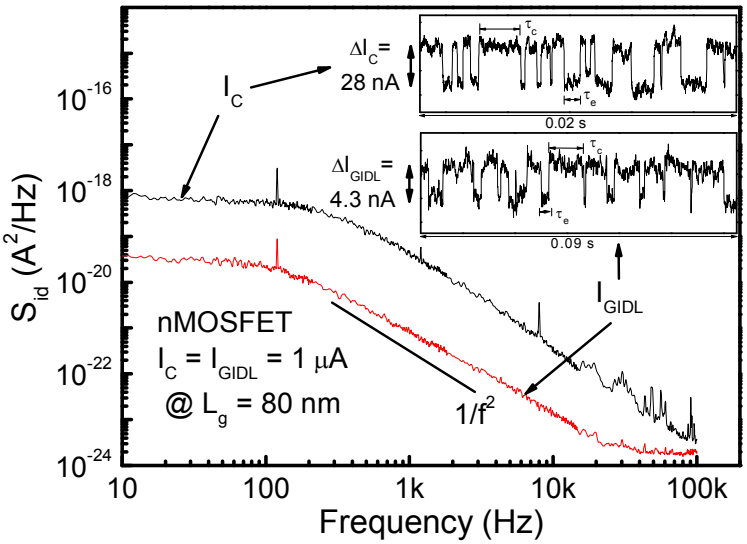
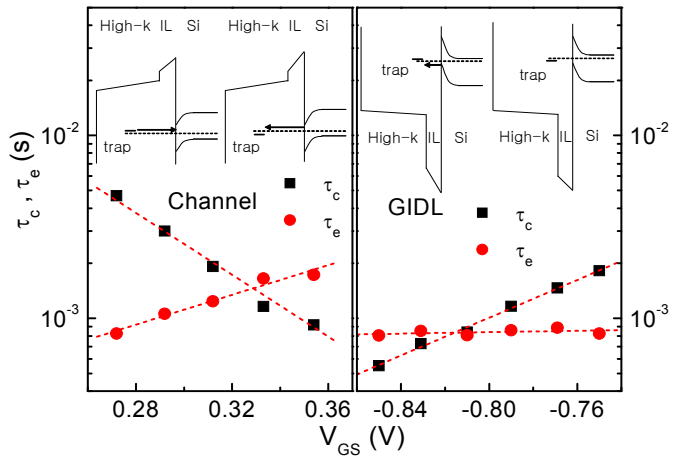


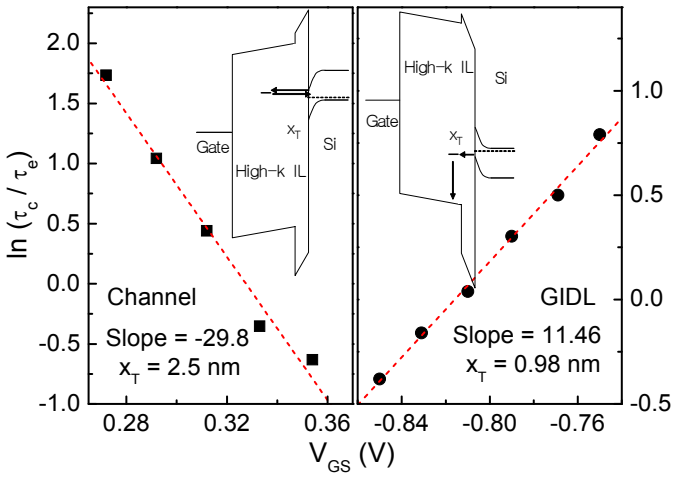
Fig. A. 6. Measured S_{id} and corresponding RTN in channel and GIDL currents of high-k nFETs measured at the same current of 1 μA.

For more investigation of LFN in GIDL currents, we characterized the LFN of GIDL currents in high-k gate dielectric nFETs [36]. High-k dielectrics, HfSiON, were deposited using an atomic layer deposition (ALD) methods after a chemical cleaning to form a SiO_2 interfacial layer. Then, TiN electrodes were deposited using an ALD. The gate stack consists of a TiN gate electrode, 2.7 nm HfSiON as the high-k layer, and a SiO_2 interfacial layer. An equivalent oxide thickness (EOT) of the interfacial layer is ~0.6 nm. We investigated the transistors having a gate width of 10 μm and a gate length of 0.06~0.5 μm. Fig. A. 6 shows measured S_{id} and corresponding RTN waveforms in I_C and I_{GIDL} currents. The upper and lower insets stand for ΔI_C and ΔI_{GIDL}, respectively,

corresponding to S_{id}s at a current of 1 μA. ΔI_C and ΔI_{GIDL} is 28 nA and 3.4 nA, respectively. Fig. A. 7 (a) shows the dependence of τ_c and τ_e on the V_{GS} in channel and GIDL currents. As V_{GS} increases, the τ_e increases and τ_c decreases in the channel current. This can be explained physically with insets as an example. As V_{GS} increases, the trap energy level (E_T) decreases as compared with the Fermi level (E_F). So, an electron can be captured more easily and an electron which is captured to a trap is emitted with lower probability. In the GIDL current, the τ_e is independent of V_{GS}, which means that a hole which is captured to a trap is thermally emitted to valence band of gate dielectric. The τ_c increases as V_{GS} increases. This can be explained physically with insets as an example. A hole can be captured more easily to a trap from the drain when the E_T is larger than E_F. But the E_T decreases as compared with the E_F with increasing V_{GS}. Therefore, the hole capture probability is decreasing with increasing V_{GS}. Fig. A. 7 (b) shows dependence of $\ln(\tau_c/\tau_e)$ on V_{GS} for channel and GIDL currents. Dashed lines represent linear fitting curves to obtain trap depth (x_T) from the interface between the drain and the oxide. The x_T can be extracted by using the conventional method, and extracted x_Ts are 2.5 nm and 0.98 nm for channel and GIDL currents, respectively. Both traps are in high-k layer. Insets represent the energy band diagrams showing trapping and de-trapping of an electron and a hole in a trap for the channel and GIDL currents, respectively.



(a)



(b)

Fig. A. 7. (a) Dependence of τ_c and τ_e on the gate voltage in channel and GIDL currents, respectively. (b) Dependence of $\ln(\tau_c/\tau_e)$ on V_{GS} for channel and GIDL currents, respectively. Dashed lines in figure (b) represent linear fitting curves to obtain the depth (x_T) of trap from the interface between the oxide and drain.

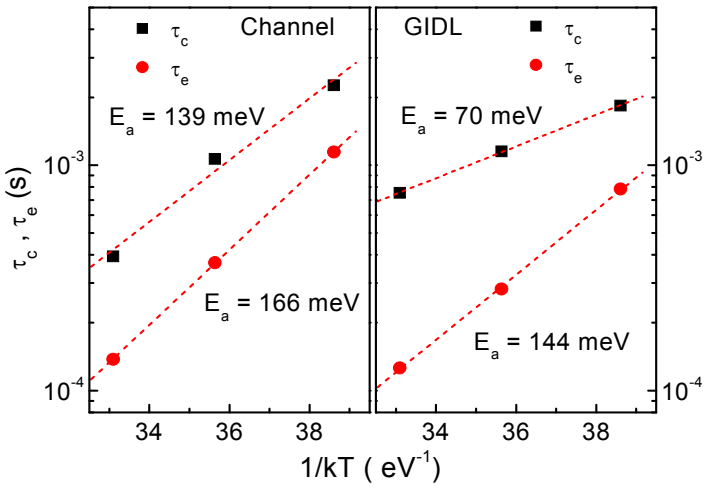


Fig. A. 8. The dependence of τ_c and τ_e on the temperature in samples shown in Fig. A. 7. The activation energies (E_a) are extracted for τ_c and τ_e of both devices.

Fig. A. 8 shows the dependence of τ_c and τ_e on the temperature in the same samples as those in Fig. A. 7. Here, all τ_c and τ_e are decreasing with increasing temperature. This means that the capture and emission events happen more frequently with increasing temperature in channel and GIDL currents because the electrons and holes can easily surpass the barrier for capture and emission. The activation energies for the τ_c and τ_e are extracted. For a trap in the channel current, the activation energy (166 meV) for an emission is slightly larger than that (139 meV) for a capture. Unlike channel current, a trap for GIDL current has much larger activation energy (144 meV) for an emission than that (70 meV) for a capture. These results mean that as temperature increases, emission process becomes much more probable in GIDL current. This fact represents that the

portion of high current state of RTN waveform in GIDL current is increasing with increasing temperature, which is really observed in measured results. The data in this figure confirm that emission process in GIDL currents is dominated by the thermal process as explained in Fig. A. 7 (a).

A.2 LFN of GIDL Current in n- and p-type MOSFETs

In this chapter, LFN of GIDL current in nano-scale *n*- and *p*-type MOSFETs are compare. The capture and emission probabilities are studied based on the capture and emission times with gate voltage and temperature. Fig. A. 9 shows measured S_{id} in GIDL currents of n and *p*MOSFETs with a L_g of 60 nm. The upper and lower insets stand for ΔI_{GIDL} of n and *p*MOSFETs, respectively, corresponding to the S_{id}s at a I_{GIDL} of 1 μA. ΔI_{GIDL}s are 1.5 nA and 2.7 nA for n and *p*MOSFETs, respectively. Average ΔI_{GIDL}s of measured 60 nm devices are 2 nA and 3 nA for n and *p*MOSFETs, respectively, and ΔI_{GIDL} of *p*MOSFETs is much scattered compared to that of *n*MOSFETs. In case of *p*MOSFETs, a trap with a captured electron which is negative charge decreases vertical electric field and decreases GIDL, which gives low current level in RTN waveforms shown in the lower inset of Fig. A. 9. High GIDL current levels are observed when a hole is emitted from a trap in an *n*MOSFET and an electron is emitted from a trap in a *p*MOSFET. In summary, RTNs in GIDL currents for n and *p*MOSFETs are originated by trapping and de-trapping of a hole and an electron, respectively.

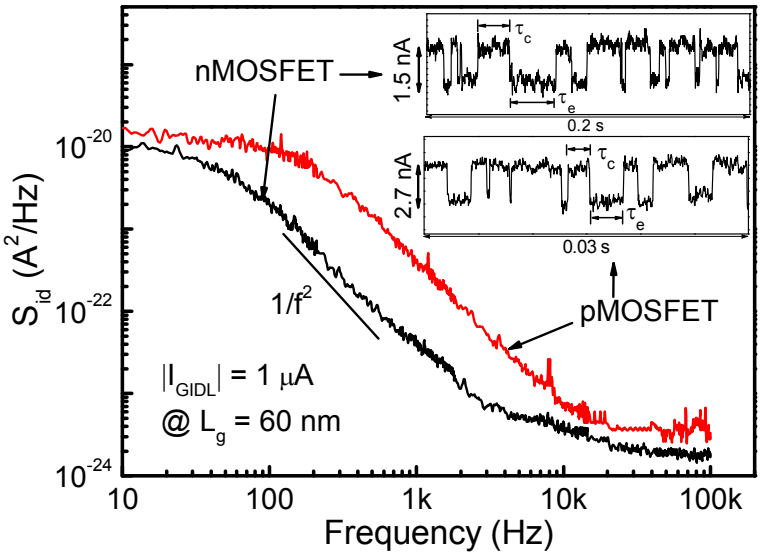
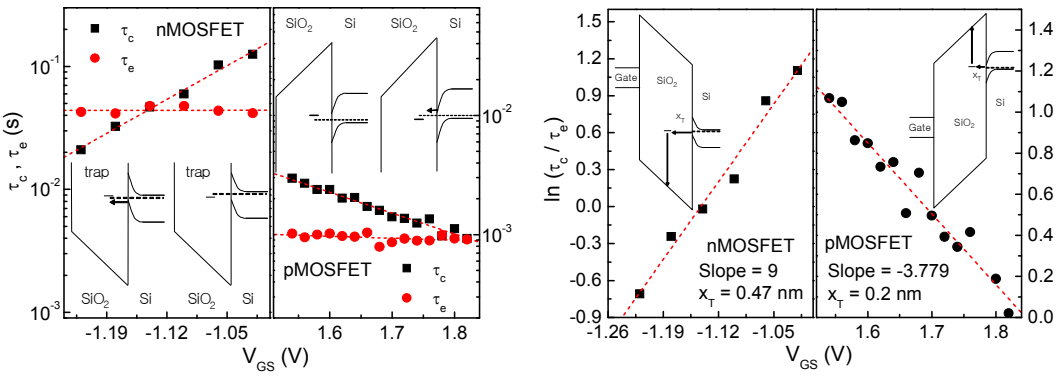


Fig. A. 9. The S_{id}s and corresponding RTN waveforms in GIDL currents of *n*- and *p*-MOSFETs measured at a fixed I_D of 1 μA.



(a) (b)

Fig. A. 10. (a) The dependence of capture time constant (τ_c) and emission time constant (τ_e) on the V_{GS} in *n* and *p*MOSFETs. (b) The dependence of ln(τ_c/τ_e) on V_{GS} for GIDL current in the *n* and *p*MOSFETs.

Fig. A. 10 (a) shows the dependence of capture time constant (τ_c) and emission time constant (τ_e) on the V_{GS} in n and *p*MOSFETs. The τ_e in an n and a *p*MOSFET is independent of V_{GS}, which means that a hole which is captured to a trap in an *n*MOSFET is thermally emitted to valence band of SiO_2 and an electron which is captured to a trap in a *p*MOSFET is thermally emitted to conduction band of SiO_2. As V_{GS} increases, the τ_c in *n*MOSFET increases and τ_c in *p*MOSFET decreases. This can be explained physically with insets as an example. As V_{GS} increases, the trap energy level (E_T) decreases as compared with the Fermi level (E_F). In an *n*MOSFET, a hole is captured more easily to a trap from the drain when the E_T is larger than E_F. Therefore, the hole capture probability is decreasing with increasing V_{GS}. In a *p*MOSFET, an electron is captured more easily to a trap from the drain when the E_F is larger than E_T. Consequently, the electron capture probability is increasing with increasing V_{GS}. Fig. A. 10 (b) shows dependence of $\ln(\tau_c/\tau_e)$ on V_{GS} for GIDL current in the n and *p*MOSFETs. Dashed lines represent linear fitting curves to obtain trap depth (x_T) from the interface between the drain and the oxide. The x_T can be extracted by using the conventional method [27], and extracted x_Ts are 0.47 nm and 0.2 nm for n and *p*MOSFETs, respectively. Insets represent the energy band diagrams showing trapping and de-trapping of a hole in a trap for the *n*MOSFET and an electron in a trap for the *p*MOSFET, respectively.

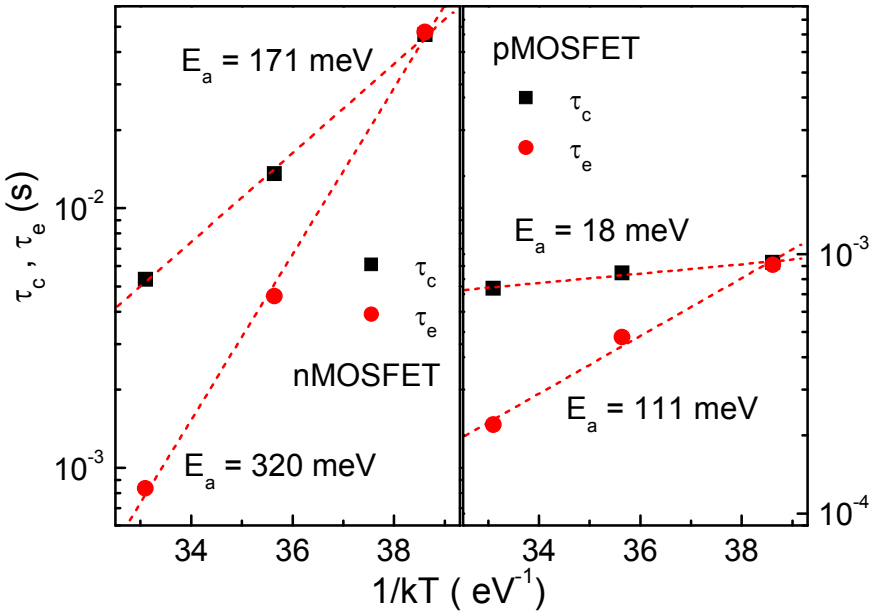


Fig. A. 11. The dependence of τ_c and τ_e on the temperature in the same samples as those in Fig. A. 10.

Fig. A. 11 shows the dependence of τ_c and τ_e on the temperature in the same samples as those in Fig. A. 10. Here, all τ_c and τ_e are decreasing with increasing temperature. This means that the capture and emission events happen more frequently with increasing temperature in both n and *p*MOSFETs because the holes and electrons can easily surpass the barrier for capture and emission. The activation energies for the τ_c and τ_e are extracted. For a trap in an *n*MOSFET, the activation energy (171 meV) for a capture is smaller than that (320 meV) for an emission. A trap in a *p*MOSFET also has larger activation energy (111 meV) for an emission than that (18 meV) for a capture. These results mean that as temperature increases, emission process becomes more probable in an *n* and a *p*MOSFET.

This fact says that the portion of high current state of RTN waveform in GIDL current is increasing with increasing temperature in both *n* and *p*MOSFETs, which is really observed in measured results. The data in this figure confirm that both emission process in *n* and *p*MOSFETs are dominated by the thermal process as explained in Fig. A. 10 (a).

The S_{id} and ΔI_D of GIDL current in many samples having various L_g (55, 60, 70, 80, 125, 200 and 400 nm) of *n* and *p*MOSFETs are measured and compared with those of channel current of *n*MOSFETs. As shown in Fig. A. 12, both S_{id} and ΔI_D of GIDL current from *n* and *p*MOSFETs are much smaller than those from channel current of *n*MOSFETs. In addition, the dependence of S_{id} and ΔI_D of GIDL current on L_g is much smaller than those of channel current and the ΔI_{ch} is much scattered compared to the ΔI_{GIDL}. These indicate that the immunity of RTN fluctuation in GIDL current is higher than channel current.

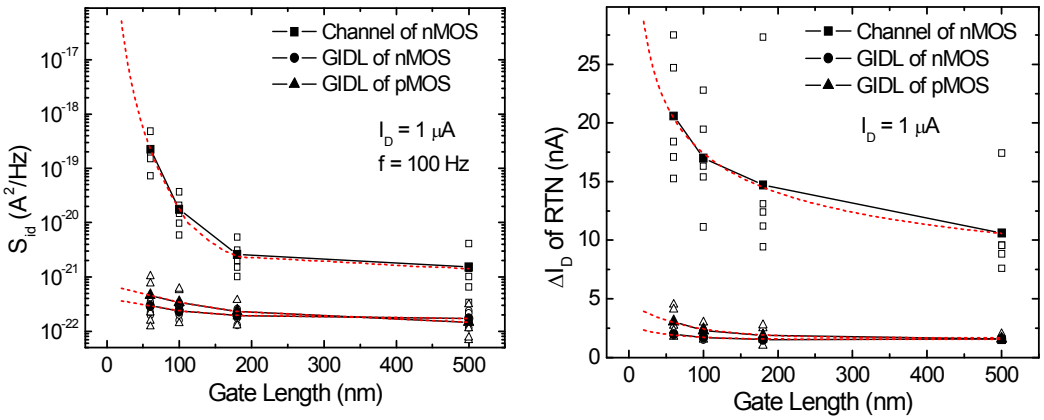
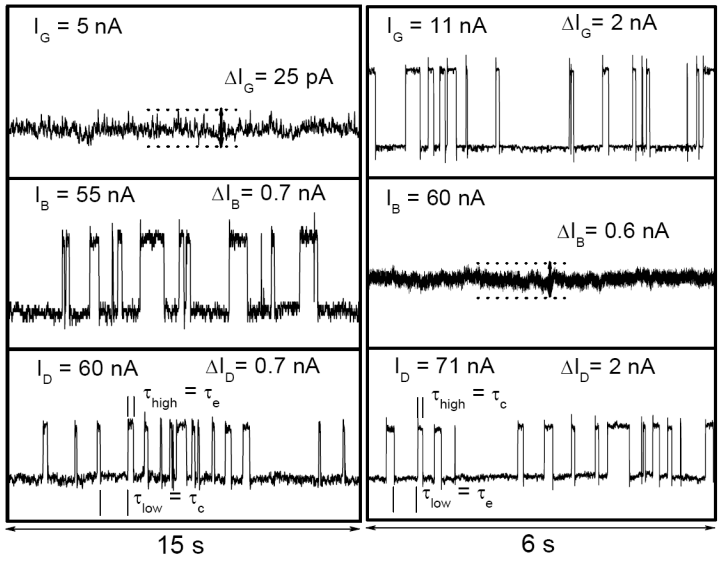


Fig. A. 12. The dependence of LFN characteristics on L_g (a) S_{id} (b) ΔI_D.

A.3 RTN in GIDL and Gate Edge Tunneling Currents

There may be a concurrent gate edge tunneling (ET) leakage current that flows between the gate and the drain (or the source) overlapped by the gate [37], [38]. The gate ET of electrons from the gate to the underlying n-type drain extension has been examined in off-state *n*MOSFETs having an ultra-thin gate oxide [38]. The gate ET is more pronounced for thin gate oxides, and gate ET affects the GIDL current measured at the drain terminal. Therefore, it is imperative to check the contribution from gate ET under gate ET bias conditions. In this chapter, the RTN in I_D, I_B and I_G of HfSiON high-k gate dielectric *n*FETs is investigated through systematic experimentation and calculation.

RTN fluctuations in I_D, I_B, and I_G were measured under GIDL bias conditions. These are measured one after the other due to the limitation for measuring three currents at the same time and plotted to the same time interval. The V_G, V_D, and V_B that were applied were −0.3 V, +1.5 V, and 0 V, respectively, for *n*FETs for the measurement of GIDL. We first checked the effect on GIDL from the n^+-p junction diode (n^+ drain to p body), and determined that the junction leakage measured at a V_D of 1.5 V and a V_B of 0 V was negligible (< 10 pA) compared to the GIDL current (> 50 nA).

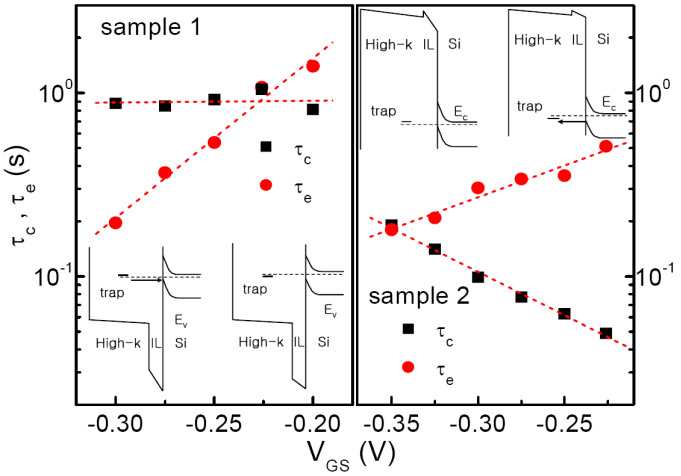


(a) (b)

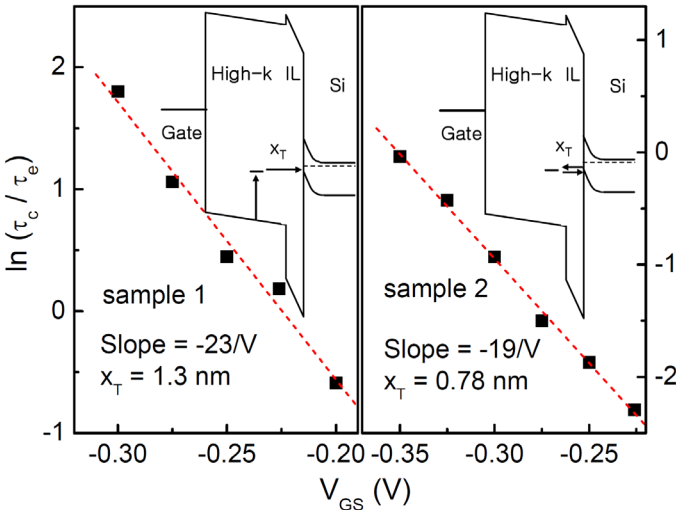
Fig. A. 13. Measured I_D, I_B and I_G values under GIDL bias conditions of V_{GS} = −0.3 V and V_{DS} = +1.5 V in *n*FETs with L_g of (a) 0.3 μm and (b) 0.08 μm. The currents are measured one after the other. The FETs in (a) and (b) show the RTN in GIDL current only and gate ET current only, respectively.

Fig. A. 13 (a) and (b) show the measured I_D, I_B, and I_G under GIDL bias conditions of V_{GS} = −0.3 V, V_{GB} = −0.3 V, and V_{DS} = +1.5 V in *n*FETs. The L_g in Fig. A. 13 (a) and (b) are 0.3 μm and 0.08 μm, respectively. The ΔI_G, ΔI_D, and ΔI_B terms represent the current fluctuations due to RTN in I_G, I_D, and I_B, respectively. I_D is equal to the sum of I_B (consisting mainly of GIDL current) and I_G (the gate ET leakage current). Most of the I_G is the result of the tunneling current which flows between the gate and the drain overlapped by the gate, because V_{GD} (−1.8 V) is much larger than V_{GS} (−0.3 V) and V_{GB}

(−0.3 V). The tunneling currents from the gate to the source and to the channel are less than 10 pA and 20 pA, respectively. In Fig. A. 13, two samples in the same wafer showed different characteristics. In Fig. A. 13 (a) (sample 1), I_D and I_B show RTN, and the ΔI_D resulting from the RTN is almost the same as ΔI_B. However, there is no RTN in I_G. Therefore, it appears that GIDL is a main cause of the RTN. Electron-hole pairs are generated in the n^+ drain overlapped by the gate; the electrons go to the drain terminal and the holes move to the body along the surface of the drain. Here, a trap with a captured electron increases the vertical electric field in a localized surface area, and increases GIDL as a result, which means that the high current level in the RTN waveform reflects an electron-captured state. Thus, a low current level in the RTN waveform can be explained by electron emission. A trap with a de-trapped electron in a gate dielectric reduces the vertical electric field, resulting in a decrease in GIDL current. Thus, the trapping and de-trapping of an electron produces RTN in GIDL current. In Fig. A. 13 (b) (sample 2), I_D and I_G show RTN, and ΔI_D is nearly the same as ΔI_G. But, there is no RTN fluctuation in I_B, which indicates that there is no active trap to generate RTN in the GIDL current. This result suggests that a trap in the gate dielectric can trap and de-trap electrons while gate ET current is flowing, which is the main cause of RTN. Here, a trap with a captured electron hinders the electron flow from the gate in a localized area and decreases I_G slightly, which leads to the low current level in RTN waveform. The de-trapping of the electron (electron-emission) results in the I_G returning to an undisturbed value which is the high current level in RTN waveform.



(a)



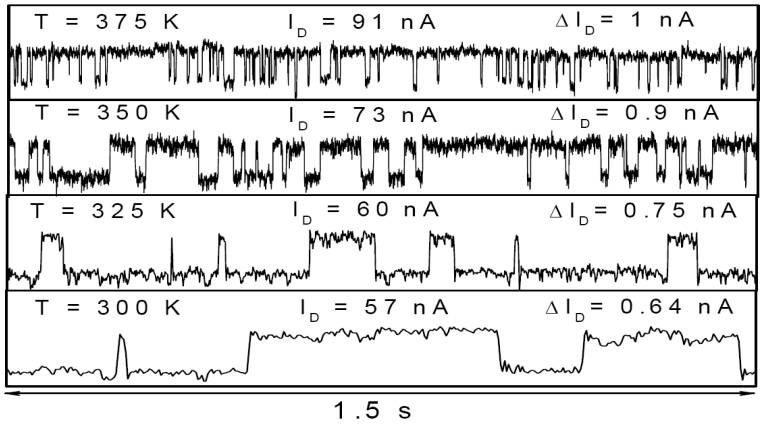
(b)

Fig. A. 14. (a) The dependence of τ_c and τ_e on the gate voltage for *n*FET samples 1 and 2. Insets present the energy band diagrams showing emission and capture probabilities versus gate voltage. (b) The dependence of $\ln(\tau_c/\tau_e)$ on V_{GS} for samples 1 and 2. Dashed lines represent linear fitted curves used to obtain x_T. Insets represent the energy band diagrams showing trapping and de-trapping of electrons for samples 1 and 2.

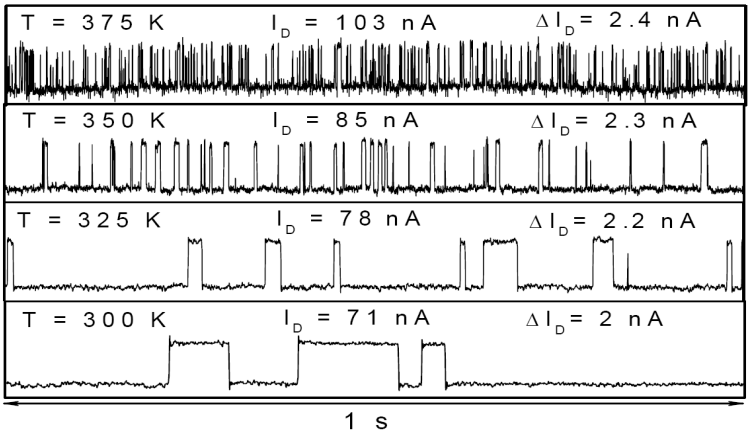
In Fig. A. 14 (a), the τ_c and τ_e on the gate voltage in samples 1 and 2 is shown. In sample 1, τ_c is independent of the gate voltage because an electron from the valence band of the high-k dielectric is thermally emitted to a de-trapped trap. However, the τ_e of sample 1 increases with increasing V_{GS}, which indicates that electron de-trapping becomes more difficult as V_{GS} increases. This can be explained using two energy band diagrams (see Fig. A. 14 insets) to depict the behavior of electron de-trapping with V_{GS}. Generally a trap energy (E_T) which is similar to Fermi energy (E_F) actively contributes to low frequency noise.

Now assume that E_T is slightly higher than E_F when the V_{GS} is low (sample 1 left inset). As V_{GS} increases, E_T becomes slightly lower than E_F (sample 1 right inset). Thus, with increasing V_{GS}, the probability of de-trapping of a captured electron decreases because of a decrease in the E_T–E_F, which results in an increasing τ_e. In sample 2, τ_e increases with increasing V_{GS}, which means that the captured electrons are mainly de-trapped; not by thermionic emission, but by the electric field. It is very difficult to emit an electron from a trap to the conduction band of a high-k dielectric because $E_{C,high\text{-}k}$–E_T is much larger than E_T–$E_{V,high\text{-}k}$ in our samples. Here, $E_{C,high\text{-}k}$ and $E_{V,high\text{-}k}$ represent the conduction band and valence band energies of the high-k dielectric, respectively. As V_{GS} increases, E_T–E_F decreases slightly as shown in the sample 2 inset, which in turn decreases the probability of electron emission from a trap to the drain. In contrast, τ_c of sample 2 decreases with increasing V_{GS}. Then E_T–E_F decreases slightly, as shown in the τ_e explanation of sample 2, which in turn, increases the capture probability of electron from

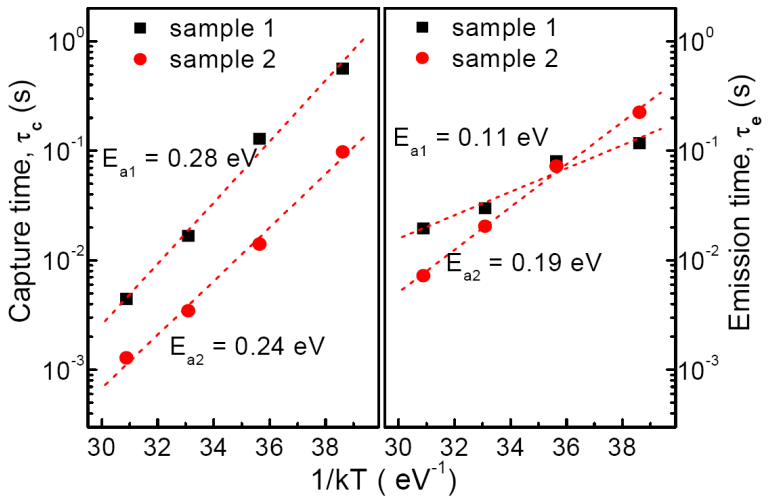
the drain to a trap. Fig. A. 14 (b) shows the $\ln(\tau_c/\tau_e)$ of the RTN in samples 1 and 2 versus V_{GS}. The trap depth (x_T) can be obtained by using a conventional method, and by considering the different dielectric constants of the SiO_2 IL and the HfSiON high-k gate dielectric. Since x_Ts for samples 1 and 2 are 1.3 nm and 0.78 nm, respectively, both traps are located in the HfSiON layer. Insets show the samples' energy band diagrams, including x_T.



(a)



(b)



(c)

Fig. A. 15. Variation of I_D-RTN waveforms with temperature at V_{GS} = −0.3 V and V_{DS} = +1.5 V over time for (a) sample 1 and (b) sample 2. (c) The dependence of τ_c and τ_e on temperature in samples 1 and 2. In (c), the activation energies are extracted.

Fig. A. 15 (a) and (b) show RTN waveforms of I_{DS}s for samples 1 and 2, respectively, in a time domain as a parameter of temperature (T). Here the same bias as that in Fig. 1 is applied to samples 1 and 2. As T increases, both I_D and ΔI_D increase because the GIDL currents are increases due to the trap-assisted generation of electron hole pairs [39] and direct tunneling gate currents are increases due to the thermionic-type of emission current [40]. Capture and emission events happen more frequently with increasing T in both samples because the electrons can easily surpass the barrier for capture and emission. It is interesting to note that the high current portion of RTN in GIDL (sample 1) increases

significantly with increasing T. As mentioned in the explanation of Fig. A. 14 (a), thermionic emission dominates the capture process in a trap concerning GIDL current. As T increases, the thermionic emission from the valence band of gate dielectric becomes easier, thus a trap captures an electron more frequently, which leads to the larger portion of high current state in sample 1. Differently in sample 2, the ratio of high and low current state is similar with T because both τ_c and τ_e are strongly dependent on V_{GS}. Fig. A. 15 (c) shows the dependence of τ_c and τ_e on T, and the activation energies extracted from these data. For a trap in sample 1, the capture activation energy (0.28 eV) is much larger than that (0.11 eV) for emission. In contrast, although the trap in sample 2 has larger capture activation energy (0.24 eV) than that (0.19 eV) for emission, the energy difference is relatively small.

Fig. A. 16 (a) and (b) show fluctuations in waveforms over time in I_D, I_B and I_G in *n*FETs under the same GIDL bias conditions. These are measured one after the other due to the limitation for measuring three currents at the same time and plotted to the same time interval. L_gs in Fig. A. 16 (a) and (b) are 100 nm and 75 nm, respectively. The device (sample 3) in Fig. A. 16 (a) has RTN in both GIDL and EDT currents because I_G and I_B exhibit two-level RTN. Since I_D consists of the sum of I_G and I_B, we can expect a four-level RTN in I_D. However, a three-level RTN pattern is depicted in I_D. The ΔI_G is nearly the same with ΔI_B at the V_{GS} of −0.3 V, which is why the three-level RTN is observed. If we change V_{GS}, then the ΔIs for both currents would be different; resulting in a four-level RTN. Note that the RTN in I_B changed more frequently than that in I_G, and that the RTN

in I_D reflects the waveforms in I_G and I_B.

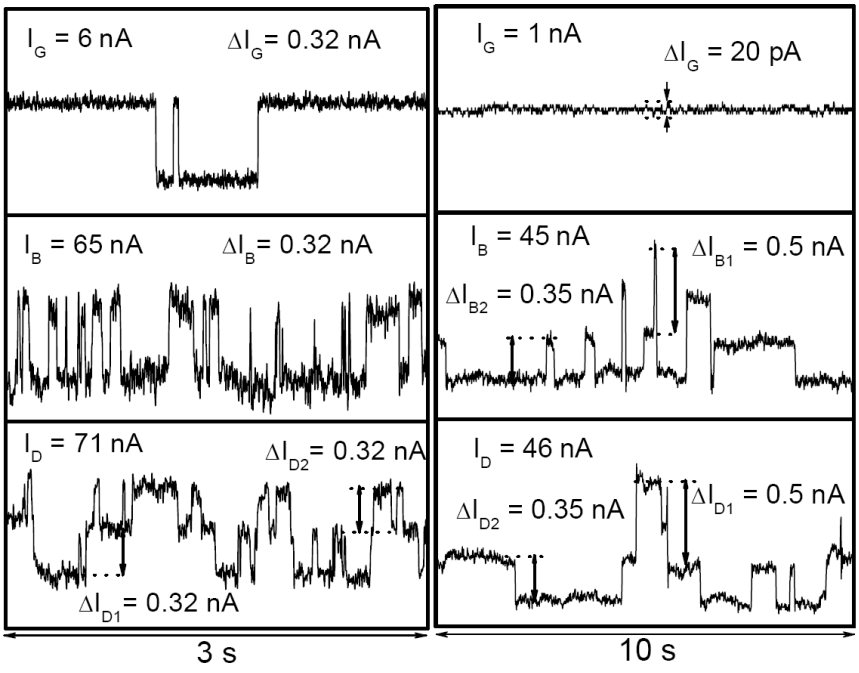
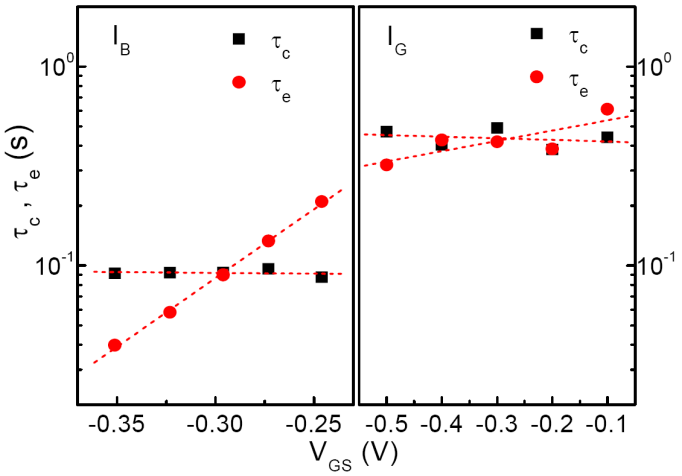


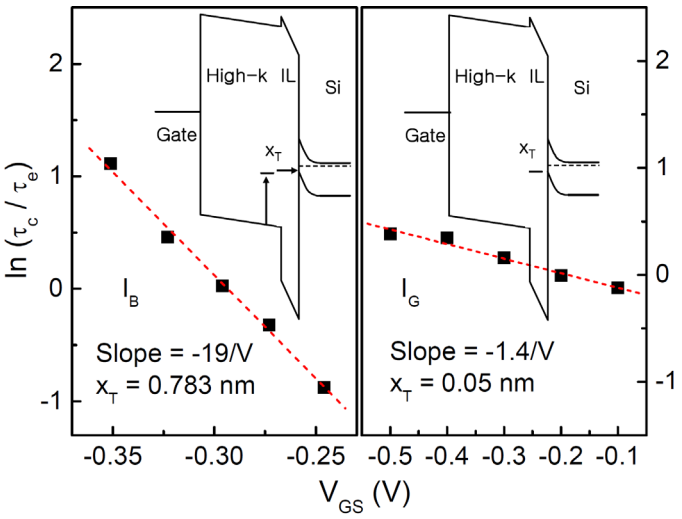
Fig. A. 16. Measured I_D, I_B and I_G values under GIDL bias conditions of V_{GS} = −0.3 V and V_{DS} = +1.5 V in *n*FETs with L_g of (a) 0.1 μm and (b) 0.075 μm. The currents are measured one after the other. The FET in (a) shows two-level RTN in both GIDL and gate ET currents while the FET in (b) shows four-level RTN in the GIDL current only.

In Fig. A. 16 (b), the *n*FET (sample 4) shows four-levels of RTN in both I_B and I_D, but no RTN in I_G. Because there is no RTN in I_G, there is no trap for the RTN in the gate ET current. The four-level RTN in sample 4 is attributed to the presence of two

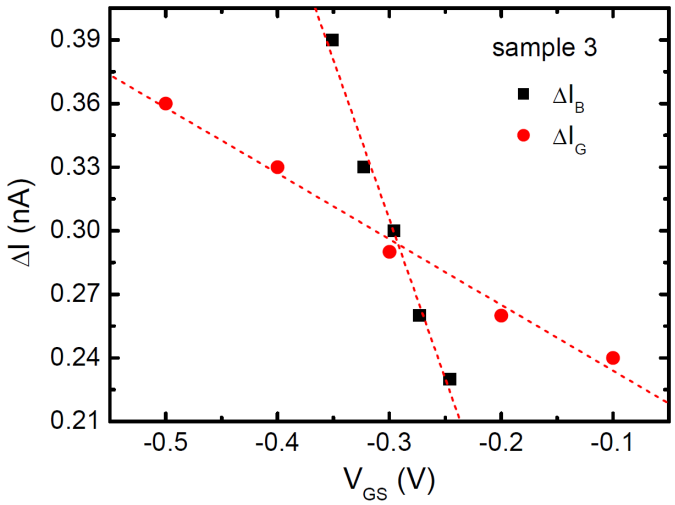
independent traps that affect the GIDL current. The ΔIs in I_D are nearly the same with those in I_B. In both I_D and I_B, the ΔI of one trap (trap 1; ΔI = 0.5 nA) is larger than that of the other trap (trap 2; ΔI = 0.35 nA). The complex RTN observed in Fig. A. 16 was analyzed by examining sample 3 in more detail.



(a)



(b)



(c)

Fig. A. 17. (a) The dependence of τ_c and τ_e on the gate voltage for I_B and I_G in sample 3. Behaviors of time constants for RTNs in I_B and I_G are same with those samples 1 and 2, respectively, shown in Fig. 2 (a). (b) The dependence of $\ln(\tau_c/\tau_e)$ on V_{GS} for the RTNs in I_B and I_G in sample 3. Dashed lines represent linear fitted curves used to obtain x_T. Insets present energy band diagrams showing trapping and de-trapping of an electron for the RTNs in I_B and I_G. (c) (ΔI)s for RTNs of GIDL and gate ET currents versus V_{GS} in sample 3

Fig. A. 17 (a) shows the dependence of τ_c and τ_e on V_{GS} for the RTNs in I_B and I_G of sample 3. For the RTN of I_B (the mainly GIDL current) of sample 3, τ_c is independent of the V_{GS} as shown in the left panel of Fig. A. 17 (a); thus indicating that an electron is captured to the trap by thermionic emission as was also indicated in the left panel of Fig.

2 (a). Also, the left panel of Fig. A. 17 (a) shows that τ_e increases with increasing V_{GS}; such behavior can be explained by the same rationale as that used for sample 1. The right panel of Fig. A. 17 (a) shows that the behavior of τ_c and τ_e with V_{GS} for the RTN of the gate ET current (I_G) is similar to that in sample 2 in which RTN is only present in the gate ET current. Fig. A. 17 (b) shows the $\ln(\tau_c/\tau_e)$ versus V_{GS} for the RTNs in I_B and I_G of sample 3. The x_Ts for the traps were extracted and the results show that the trap responsible for the RTN in GIDL is located in high-k dielectric layer (x_T = 0.783 nm). The IL includes the trap (x_T = 0.05 nm) which affects the RTN in the gate ET current. The Fig. 5 (b) insets show energy band diagrams, including x_T. From the results of sample 3 in Fig. A. 16 (a), which shows RTNs in GIDL and gate ET currents, we understand that the RTN in I_D comes from a combination of the RTNs in GIDL and gate ET currents. Fig. A. 17 (c) shows ΔIs in the RTNs of the GIDL and gate ET currents versus V_{GS}. At a V_{GS} of −0.3 V, both RTNs are similar, which produces a three-level RTN in I_D as was also shown in Fig. A. 17 (a). As the V_{GS} increases, ΔI of the GIDL current decreases more rapidly compared to the ΔI of the gate ET current. At V_{GS}s values other than −0.3 V, ΔIs in both RTNs are different, which results in a four-level RTN.

Fig. A. 18 shows $\ln(\tau_c/\tau_e)$ versus V_{GS} for the RTNs in I_B or I_D of sample 4 which had two independent traps that were responsible for producing the four-level RTN in the GIDL current. The behavior of τ_c and τ_e versus V_{GS} in the RTN for each trap is quite similar to that for the RTN in the GIDL current shown in Fig. A. 14 (a). The extracted x_Ts for traps 1 and 2 are 0.55 nm and 1.164 nm, respectively. Fig. A. 18 inset shows

schematically the vertical positions of traps; and traps 1 and 2 are located in IL and high-k layer, respectively

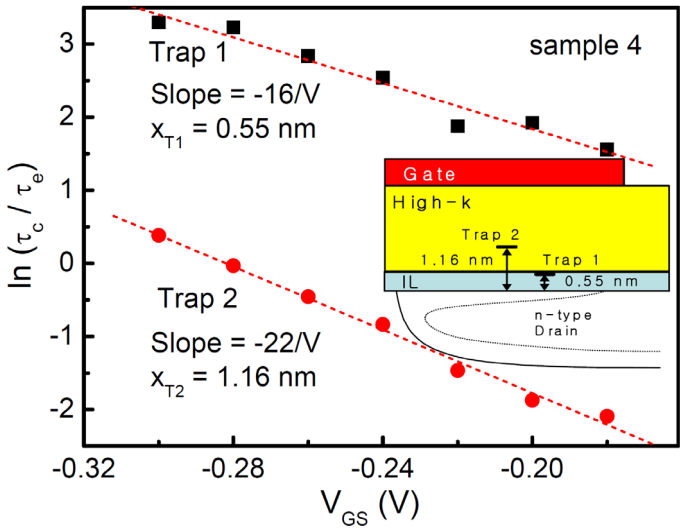
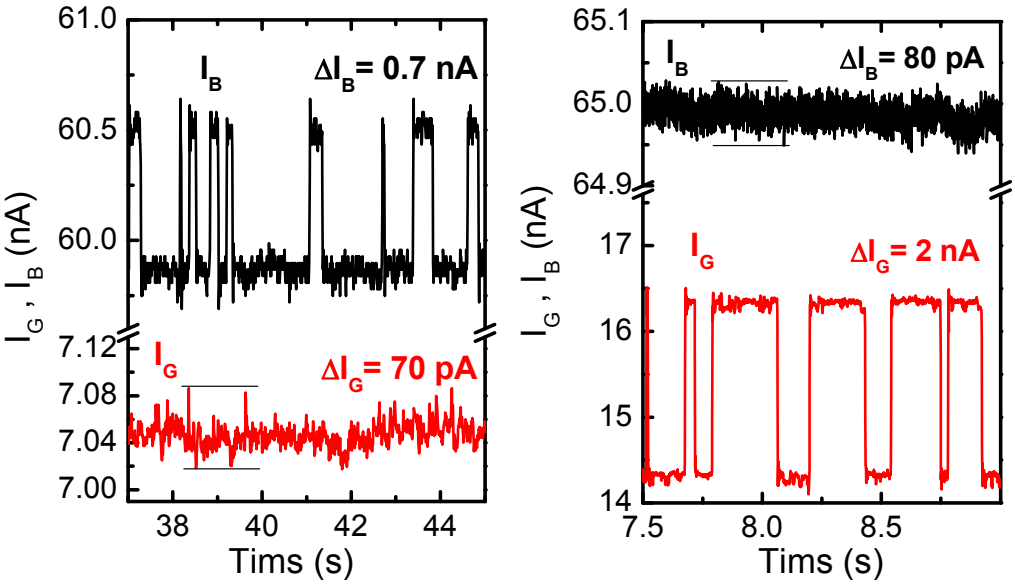


Fig. A. 18. The dependence of ln(τ_c/τ_e) on V_{GS} for traps 1 and 2 in sample 4. Dashed lines represent linear fitted curves used to obtain x_T. Traps 1 and 2 are located in the IL and HfSiON high-k layer, respectively. The inset schematically indicates the vertical positions of the two traps.

A.4 Model for RTN in GIDL Current

In this chapter, model of RTN in GIDL current was proposed and explained together with that of gate ET current. The RTNs in the gate ET current as well as the GIDL current were observed under GIDL bias condition in high-k *n*MOSFETs [41].The relative amplitude of the RTN (ΔI/I) in the gate ET current was extracted and a model for elucidating the generation of the RTN in the current was established [42]. Both I_{GIDL} and gate ET currents (I_{ET}) are flown from the drain overlapped by the gate. So, it is expected that a trap inside the gate dielectric between the drain and the gate can affect I_{GIDL} and I_{ET}, simultaneously. To examine the synchronism between RTNs of I_{GIDL} and I_{ET}, it is essential to measure these two currents at the same time. Fig. A. 19 (a) and (b) show simultaneously measured gate and body currents (I_G and I_B) in the devices 1 and 2, respectively, which have the RTN in the I_B or I_G under GIDL bias conditions of V_{GS} = −0.3 V and V_{DS} = 1.5 V. Although the drain and source currents (I_D and I_S) are not shown, the I_D is the sum of I_G and I_B, and the I_S is negligible (< 10 pA). So, the I_G is nearly the same as the I_{ET} and the I_B is nearly the same as the I_{GIDL}. Therefore, the RTNs in devices 1 and 2 are generated in the I_{GIDL} and the I_{ET}, respectively. As shown in Fig. 1, there is no correlation between the I_B and the I_G. This means that a trap which induces the RTN in the I_{ET} has no effect on the I_{GIDL} and a trap which induces the RTN in the I_{GIDL} does not influence on the I_{ET}. Now we investigate the reason why the RTNs in the I_{GIDL} and I_{ET}

have no correlation in these devices



(a) (b)

Fig. A. 19. Simultaneously measured gate and body currents (I_G and I_B) in the two devices showing the RTN in the (a) [device 1] GIDL current and (b) [device 2] gate ET current under GIDL bias condition of V_{GS} = −0.3 V and V_{DS} = 1.5 V.

Fig. A. 20 shows a schematic cross sectional view showing three different traps in the drain overlapped by the gate. There are three cases corresponding to three traps. Case 1 depicts the situation that there is no defect region in several Debye lengths from a trap located in the gate dielectric. In case 2, an I_{ET} percolation path through the dielectric is located near a trap. Case 3 delineates that an I_{GIDL} percolation path is located near a trap

inside the dielectric. We performed 3-D device simulation with these cases using the SILVACO tool to investigate I_{GIDL} change with the capture and emission of an electron. The simulated device structure is made equal to that of measured device. The channel width is 10 μm and the overlap length of the drain overlapped by the gate is 5 nm. In case 1, a trap is located at an arbitrary position inside the gate dielectric on the drain overlapped by the gate. Then, we investigate the difference between the I_{GIDL}s when the trap is neutral and charged with −1q (-1.6×10^{-19} C) under GIDL bias condition (V_{GS}= −0.3 V and V_{DS}= 1.5 V). As a result, the change of the I_{GIDL} (ΔI_{GIDL}) was ~7 pA which is extremely small compared with the measured one. Although the position of the trap was changed vertically and laterally, the results were similar. A trap in case 1 cannot make appreciable RTN shown in the Fig. A. 19 (a).

Case 2 shows that a schematic model for elucidating the generation of the RTN in the I_{ET} [42]. We also performed device simulation using a silicon pillar with a modified mobility to mimic a percolation path and a trap located near the percolation path. As a result, the big difference between the I_{ET}s when the trap is neutral and charged with -1q could be observed. But the ΔI_{GIDL} was still very small in this case. From this model, the reason why a trap which induces appreciable RTN in the I_{GIDL} has no effect on the I_{ET} can be explained as follows. An I_{ET} percolation path should not exist in several Debye lengths from the trap which induces appreciable RTN in the I_{GIDL}. Therefore, the I_{ET} fluctuation due to a trap which induces the RTN in the I_{GIDL} is ignorable.

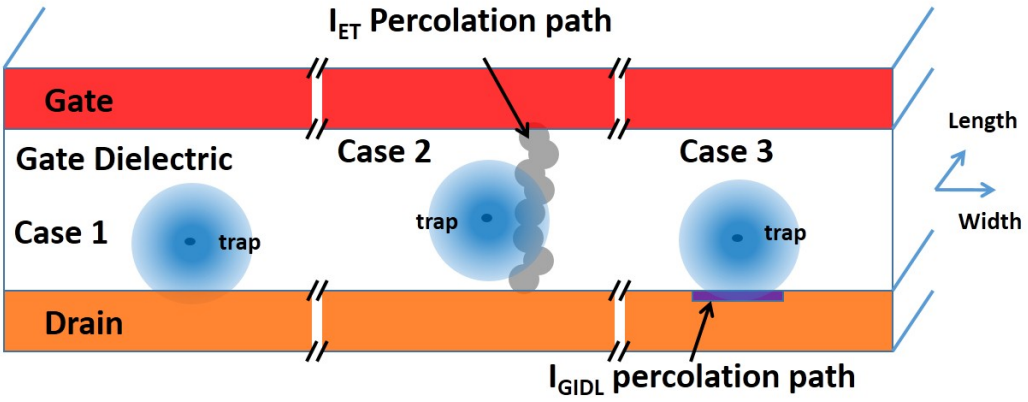


Fig. A. 20. A schematic cross sectional view showing three different traps located in the gate dielectric on the drain overlapped by the gate along the width direction.

Now we have two questions unsolved: One is the physical mechanism for the generation of appreciable RTN in the I_{GIDL} and the other is the reason why a trap which induces the RTN in the I_{ET} has no influence on the I_{GIDL}. In order to answer these two questions, we propose a physical model shown in case 3 of Fig. A. 20. To be answered physically, a localized defect region (called as I_{GIDL} percolation path) should exist on the surface of the drain region and should be located near a trap. The defect region has short lifetimes of carriers and gives much higher I_{GIDL} than those of normal drain region. If a trap aligns with the defect region, appreciable RTN in the I_{GIDL} as shown in device 1 of Fig. A. 19 (a) can be observed. In order to confirm this model, we performed the device simulation for case 3. A small region (3 nm × 3 nm) for the defect region is formed with a depth of 1 nm at the surface of the drain overlapped by the gate in the simulated device. In the small region, two constants of band-to-band generation model (Klaassen model)

are modified to mimic a localized defect region. The increase of GIDL current due to the percolation path is ~11.4% which is within the range of the variation (maximum variation: ~42%) of GIDL current measured in several tens devices having the same device size. A trap is located inside the gate dielectric on the defect region. Through the device simulation, we obtained a reasonable ΔI_{GIDL} (~1 nA) which is comparable with experimental one. Thus the requirement for the generation of appreciable RTN in the I_{GIDL} is that a trap should be located near an I_{GIDL} percolation region, which is the answer for question 1. The answer for question 2 is that an I_{GIDL} percolation path should not exist in several Debye lengths from the trap which induces appreciable RTN in the I_{ET}. So, the I_{GIDL} fluctuation due to the trap which induces the RTN in the I_{ET} is ignorable.

From these two models (cases 2 and 3) shown in Fig. A. 20, we can explain physically the mechanism for the generation of the RTNs in the I_{GIDL} and I_{ET}, and no correlation between the RTNs in both currents shown in Fig. A. 19. In the previous study, the RTNs were observed simultaneously in both I_{GIDL} and I_{ET} in a device although both RTNs were not synchronized. This device has uncorrelated I_{ET} and I_{GIDL} percolation paths, which means that they were far away. If both percolation paths are located closely, a trap located both percolation paths can generate synchronized RTNs in both I_{GIDL} and I_{ET}. However, synchronized RTNs in both currents in all devices we have measured have not been observed since this probability is extremely small.

Bibliography

[1] R. Bez, E. camerlenghi, A. Modelli, and A. Visconi, “Introduction to Flash Memory,” *Proc IEEE*, Vol. 91, pp.489-502, Apr. 2003.

[2] Kawamatus, Tatsuya. "TECHNOLOGY FOR MANAGING NAND FLASH". Hagiwara sys-com co., LTD., August 1, 2011.

[3] R. Micheloni, L. Crippa, A. Marelli, “Inside NAND Flash Memories”, *Springer*, 2010.

[4] S. Tam, P. K. Ko, and C. Hu, “Lucky-Electron Model of Channel Hot-Electron Injection in MOSFET’s,” *IEEE Transactions on Electron Devices*, vol. 31, no. 9, pp.1116-1125, Sept. 1984.

[5] Sumio Tanaka, and Mitsuaki Ishikawa, “One-Dimensional Writing Model of n-Channel Floating Gate Ionization-Injection MOS (FIMOS),” *IEEE Transactions on Electron Devices*, vol. 28, no. 10, pp.1190-1197, Oct. 1981.

[6] C.H. Lee et al., “Highly scalable NAND Flash memory with robust immunity to program disturbance using symmetric inversion-type source and drain structure”, *Symposium on VLSI technology,* pp. 118-119, 17-19 June 2008.

[7] Eun Suk Cho, Hyun Jung Kim, Byoung Taek Kim, Jai Hyuk Song, Du Heon Song, Jeong-Hyuk Choi, Kang-Deog Suh, and Chilhee Chung, “Optimal Cell Design for

Enhancing Reliability Characteristics for sub 30 nm NAND Flash Memory," in *IEEE IRPS Tech. Dig.*, pp.611-614, May 2010.

[8] J. Wileu, T.Y. Fu, T Tanaka, et al., "Phase-shift mask pattern accuracy requirements and inspection technology," *Proc. SPIE 1464, Integrated Circuit Metrology, Inspection, and Process Control V, 346,* Jul., 1991.

[9] Mario Garza, Nicholas K. Eib, Keith K. Chao, et al., "Optical Proximity Correction Method and Apparatus," U.S. Patent No. 5,723,233, Mar. 3, 1998.

[10] Soichi Owa, Hiroyuki Nagasaka, "Immersion lithography; its potential performance and issues," *Proc. SPIE 5040, Optical Microlithography XVI, 724,* Jun. 2003.

[11] Yohwan Koh, "NAND Flash Scaling Beyond 20nm," *IEEE International Memory Workshop,* pp.1-3, 10-14 May 2009.

[12] H. Tanaka et al., "Bit Cost Scalable technology with punch and plug process for ultra high density Flash memory", *Symposium on VLSI technology,* pp.14-15, 2007.

[13] J. Jang et al., "Vertical cell array using TCAT (Terabit Cell Array Transistor) technology for ultra high density NAND Flash memory", *Symposium on VLSI technology,* pp.192-193, 2009.

[14] W. Kim, S. Choi, J. Sung, T. Lee, C. Park, H. Ko, J. Jung, I. Yoo, and Y. Park, "Multi-layered vertical gate NAND flash overcoming stacking limit for terabit density storage," in *VLSI Symp. Tech. Dig.,* pp. 188–189, 2009.

[15] R. Waser and M. Aono "Nanoionics-based Resistive Switching Memories," in *Nature Materials*. vol. 6 p. 833, 2007.

[16] S. C. Chae, J. S. Lee, S. Kim, S. B. Lee, S. H. Chang, C. Liu, B. Kahng, H. Shin, D.-W. Kim, C. U. Jung, S. Seo, M.-J. Lee, and T. W. Noh in *Adv. Mater*. vol. 20, p. 1154, 2008.

[17] M. Terai, S. Kotsuji, H. Hada, N. Iguchi, T. Ichihashi, and S. Fujieda *IRPS Tech. Dig*., p. 134. 2009.

[18] C. Cagli, D. Ielmini, F. Nardi, and A. L. Lacaita *IEDM Tech. Dig*., p. 301, 2008.

[19] K. F. You and C. Y. Wu, "A new quasi-2-D model for hot-carrier band-to-band tunneling current," *IEEE Trans. Electron Devices*, vol. 46, p. 1174, June 1999.

[20] K. Roy and S. C. Prasad, Low-Power CMOS VLSI Circuit Design. New York: Wiley, ch. 2, pp. 28–29, 2000.

[21] Il Han Park, Wook-Ghee Hahn, Ki-Whan Song, Ki Hwan Choi, Hyun-Ki Choi, Sung Bok Lee, Chang-Sub Lee, Jai Hyuk Song, Jin Man Han, Kye Hyun Kyoung, Young-Hyun Jun, "A New GIDL Phenomenon by Field Effect of Neighboring Cell Transistors and Its Control Solutions in Sub-30 nm NAND Flash Devices" in *VLSI Symp*. *Tech*. *Dig*., p. 23, 2012.

[22] Eiji Yoshida and Tetsu Tanaka, "A Design of a Capacitorless IT-DRAM Cell Using Gate-induced Drain Leakage (GIDL) Current for Low-power and High-speed Embedded Memory" in *IEDM Tech. Dig*., p. 913, 2003.

[23] Alvaro Padilla, Sunyeong Lee, David Carlton, and Tsu-Jae King Liu, "Enhanced Endurance of Dual-bit SONOS NVM Cells using the GIDL Read Method" in *VLSI Symp*. *Tech*. *Dig*., p. 142, 2008.

[24] W.J. Tsai, T.F. Ou, H.L. Kao, E.K. Lai+, Y.Y. Liao, C.C. Yeh, Tahui Wang*, Joseph Ku, and Chih-Yuan Lu, “A Novel Non-volatile memory cell using a gated-diode structure with a trapping-nitride storage layer,” *VLSI Symp*. *Tech*. *Dig*., 2006, p. 42.

[25] Wen-Jer Tsai, Tien-Fan Ou, Hsuan-Ling Kao, Erh-Kun Lai, Jyun-Siang Huang, Lit-Ho Chong, Yi-Ying Liao, Shih-Ping Hong, Ming-Tsung Wu, Shih-Chang Tsai, Chia-Hao Leng, Fang-Hao Hsu, Szu-Yu Wang, Chun-Ming Cheng, Tuung Luoh, Yung-Tai Hung, Shing-Ann Luo, Chih-Hao Huang, Tao-Cheng Lu, Tahone Yang, Kuang-Chao Chen, and Chih-Yuan Lu, “A Novel Trapping-Nitride-Storage Non-Volatile Memory Cell Using a Gated-Diode Structure With an Ultra-Thin Dielectric Dopant Diffusion Barrier,” *IEEE Transactions on Electron Devices*, vol. 55, p. 2202, 2008.

[26] M. J. Kirton, M. J. Uren, “Noise in solid-state microstructures: A new perspective on individual defects, interface states and low-frequency (1/f) noise,” *Adv. Phys*., vol. 38, no. 4, p. 367, 1989.

[27] K. K. Hung, P. K. Ko, C. Hu, Y. C. Cheng, “Random telegraph noise of deep-submicrometer MOSFET’s,” *IEEE Electron Device Lett*., vol. 11, no. 2, p. 90, 1990.

[28] S. Kobayashi, M. Saitoh, K. Uchida, “Id fluctuation by stochastic single-hole trapping in high-k dielectric p-MOSFETs,” in *VLSI Symp*. *Tech*. *Dig*., 2008, p. 78.

[29] J. H. Lee, S. Y. Kim, I. H. Cho, S. B. Hwang, J. H. Lee, “1/f noise characteristics of sub-100 nm MOS transistors,” in *Journal of Semiconductor Technology and Science*, vol. 6, no. 1, 2006, p. 38.

[30] K. Fukuda, Y. Shimizu, K. Amemiya, M. Kamoshida, H. C. Hu, “Random telegraph

noise in flash memories – model and technology scaling," in *IEDM Tech. Dig*., 2007, p.169.

[31] N. Tega, H. Miki, F. Pagette, D. J. Frank, A. Ray, M. J. Rooks, W. Haensch, K. Torii, "Increasing threshold voltage variation due to random telegraph noise in FETs as gate lengths scale to 20 nm," in *VLSI Symp. Tech. Dig*., 2009, p. 50.

K. Saino, S. Horiba, S. Uchiyama, Y. Takaishi, M. Takenaka, T. Uchida, Y. Takada, K. Koyama, H. Miyake, C. Hu, "Impact of gate-induced drain leakage current on the tail distribution of DRAM data retention time," in *IEDM Tech. Dig*., 2000, p.837.

[32] K. Saino, S. Horiba, S. Uchiyama, Y. Takaishi, M. Takenaka, T. Uchida, Y. Takada, K. Koyama, H. Miyake, C. Hu, "Impact of gate-induced drain leakage current on the tail distribution of DRAM data retention time," in *IEDM Tech. Dig*., 2000, p.837.

[33] H. Kim, K. Kim, T. K. Oh, S. Y. Cha, S. J. Hong, S. W. Park, H. Shin, "RTS-like fluctuation in gate induced drain leakage current of Saddle-Fin type DRAM cell transistor," in *IEDM Tech. Dig*., 2009, p.271.

[34] A. Padilla, S. Y. Lee, D, Carlton, T. J. King, "Enhanced endurance of dual-bit SONOS NVM cells using the GIDL read method," in *VLSI Symp. Tech. Dig*., 2008, p.142.

[35] T. Krishnamohan, D. H. Kim, S. Raghunathan, K. Saraswat, "Double-gate strained-Ge heterostructure tunneling FET (TFET) with record high drive currents and <60mV/dec subthreshold slope," in *IEDM Tech. Dig*., 2008, p.947.

[36] P. D. Kirsch, M. A. Quevedo-Lopez, S. A. Krishnan, B. H. Lee, "Mobility and charge trapping comparison for crystalline and amorphous HfON and HfSiON gate dielectrics,"

Appl. Phys. Lett, 2006, 89, 242909.

[37] N. Yang, W. K. Henson, J. J. Wortman, "Analysis of tunneling currents and reliability of NMOSFET's with sub-2 nm gate oxides," *IEDM Tech. Dig*. 1999, 453.

[38] K. N. Yang, H. T. Huang, M. J. Chen, Y. M. Lin, M. C. Yu, S. M. Jang, C. H. Douglas, M. S. Liang, "Characterization and modeling of edge direct tunneling (EDT) leakage in ultrathin gate oxide MOSFETs," *IEEE Trans. Electron Device*. 2001, 48, 6, 1159.

[39] M. Rosar, B. Leroy and G. Schweeger, "A new model for the description of gate voltage and temperature dependence of gate induced drain leakage (GIDL) in the low electric field region," *IEEE Trans. Electron Device*. 2000, 47, 1, 154.

[40] A. Yassine and R. Hijab, "Temperature dependence of gate current in ultra thin SiO2 in direct-tunneling regime," *IEEE International Integrated Reliability Workshop*. 1997, p.56.

[41] J. W. Lee, B. H. Lee, H. Shin, and J. H. Lee, "Investigation of Random Telegraph Noise in Gate Induced Drain Leakage (GIDL) and Gate Edge Direct Tunneling (EDT) Currents of High-k MOSFETs," *IEEE Transactions on Electron Devices*, vol. 57, p. 913, 2010.

[42] J. W. Lee, C. H. Park, H. Shin, B. G. Park, and J. H. Lee, "Accurate Extraction of $\Delta I/I$ Due to RTN in Gate Edge Current of High-k nMOSFETs under Accumulation Mode," *Appl. Phys. Lett*. Vol. 98, p. 023505, 2011.

초 록

본 논문에서는 차세대 초고집적 메모리 소자를 위해 GIDL 전류를 센싱 전류로 이용하는 게이티드 다이오드 메모리 셀 및 어레이에 관한 연구를 진행하였다. 기존의 채널 전류를 센싱 전류로 이용하는 메모리에서, 메모리의 집적도를 높이기 위해 채널 길이를 줄이면서 발생되는 short-channel effect (SCE)가 큰 문제가 되고 있다. 하지만 게이티드 다이오드 메모리에서는 SCE 없이 소자의 크기를 줄일 수 있어서 고집적에 유리하다. 또한 기존의 낸드 플래시 메모리 어레이와 유사한 구조이면서 random access가 가능한 장점을 가질 수 있다. 게이티드 다이오드 메모리에서는, Fowler-Nordheim (FN) 터널링을 이용하여 전자를 주입하고 band-to-band 터널링에 의한 hot-hole을 이용하여 정공을 주입하여 센싱 전류인 GIDL 전류를 변화시키고 이 변화를 센싱하여 메모리 동작을 가능하게 한다.

1장에서는 최근 비휘발성 메모리의 동향에 대해 소개하였다. 2장에서는 GIDL 전류와 이를 이용하는 게이티드 다이오드 메모리에 대해 소개하였다. 3장에서는 GIDL전류를 센싱 전류로 이용하는 게이티드 다이오드 메모리의 제작을 위한 공정과정과 제작된 셀 및 어레이의 메모리 동작을 측정을 통해 검증하였다. 마지막으로 센싱 전류인 GIDL 전류의 증가를 위해 작은 밴드갭을 가지는 SiGe을 이용하는 구조를 소개하였고 시뮬레이션을 통해

향상된 소자 특성을 검증하였다. 부록에서는 GIDL 전류의 안정성 조사를 위해 MOSFET에서 GIDL 전류의 저주파 잡음을 조사하였다.

주요어: 비휘발성 메모리, 게이티드 다이오드, 밴드투밴드 터널링, GIDL, 저주파 잡음

학 번: 2009-30927